EUI DEPARTMENT OF ECONOMICS

Understanding Migration and Education Choices under Uncertainty

Alaitz Ayarza-Astigarraga

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 23 May 2023

European University Institute
Department of Economics

Understanding Migration and Education Choices under Uncertainty

Alaitz Ayarza-Astigarraga

Thesis submitted for assessment with a view to
obtaining the degree of Doctor of Economics
of the European University Institute

Examining Board

Prof. Sule Alan, EUI, Supervisor
Prof. Michèle Belot, Cornell University, Co-Supervisor
Prof. Raquel Carrasco, Universidad Carlos III de Madrid
Prof. Marco Francesconi, University of Essex

Researcher declaration to accompany the submission of written work
Department of Economics – Doctoral Programme

I, Alaitz Ayarza-Astigarraga, certify that I am the author of the work "Understanding Migration and Education Choices under Uncertainty" I have presented for examination for the Ph.D. at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

I certify that this work complies with the Code of Ethics in Academic Research issued by the European University Institute (IUE 332/2/10 (CA 297).

The copyright of this work rests with its author. Quotation from this thesis is permitted, provided that full acknowledgement is made. This work may not be reproduced without my prior written consent. This authorisation does not, to the best of my knowledge, infringe the rights of any third party.

I declare that this work consists of 25831 words.

Statement of inclusion of previous work:

I confirm that chapter 3 was jointly co-authored with Josep Amer-Mestre, Marta C. Lopes and I contributed 33% of the work.

Signature and date:

Alaitz Ayarza-Astigarraga

09/05/2023

Abstract

This thesis is composed of two related and one independent chapters. The first two chapters use rich data on subjective expectations about migration decisions of highly educated young adults from a lagging-behind region of Spain (Andalusia) that I collected myself during their migration decision-making process.

In Chapter 1, I study how expected pecuniary and nonpecuniary factors influence these individuals' migration decisions. To do so, I estimate a life-cycle model of migration choice that takes expected migration duration into account. Crucially, the collected data allow me to separate preferences from beliefs and to distinguish between pecuniary and nonpecuniary factors. Although there is sorting on expected labor market outcomes, my results show that the set of nonpecuniary factors, such as being close to family and quality of social life, play a larger role in choosing whether to migrate. Additionally, counterfactual exercises reveal that a human capital acquisition strategy has a limited effect on temporary migration plans, which are primarily driven by nonpecuniary factors.

Chapter 2 studies their self-selection intentions into internal and international migration. I find that individuals who plan to migrate internationally come from the highest end of the grade distribution and are from more privileged family backgrounds relative to the other two groups. Despite being positively selected, students who plan to migrate internationally have the most pessimistic views about their career prospects in their home region. With their migration plans, they expect higher labor market returns to migration than internal migrants. International migrants are more likely than internal migrants to plan a long term migration as opposed to a temporary migration. If individuals follow their plans, my results suggest a future brain drain from the region as well as from the country.

Chapter 3, joint with Josep Amer-Mestre and Marta C. Lopes, studies the impact of COVID-19 school closures on differences in online learning usage by regional academic performance in Italy. Using real-time data from Google Trends, we find that students in regions with a previously lower academic performance increased their searches for e-learning tools more than those with higher academic performance. Given evidence from survey and administrative

data that lower performing regions were using no less online learning before the pandemic, our results suggest that the COVID-19 shock widened the e-learning usage gap between academically lower and higher-performing regions in Italy.

Acknowledgements

This thesis would not have been possible without the support of many people. Firstly, I want to express my gratitude to my advisors, Sule Alan and Michèle Belot. Together, they formed an invaluable team of support for me, and I am very grateful to them for encouraging me to believe in my research. Most importantly, I want to thank them for their continuous guidance and for helping me formulate the right questions. I feel truly privileged for having them as role models and for getting inspired by their passion for the research they pursue.

I also want to thank Alessandro Tarozzi and Jesús Bueren for keeping their doors open and for their feedback, encouragement and support. I would also like to thank Raquel Carrasco and Marco Francesconi for generously agreeing to be part of my thesis committee.

One chapter of this thesis is coauthored with Josep and Marta. I want to thank them for their support and generosity, and for all what I learned while writing the paper together.

My time in the department would not have been the same without the amazing colleagues with whom I started the journey and that have now become great friends. Dalila, Gabriele, Josep, Konuray, Lukas, Miguel, Pietro, Thomas and Zheng, thank you for helping me be a better researcher while having fun together. I am excited to see how this little family spreads around the world and see us thrive in our careers. I am also thankful to the rest of my cohort and to the bigger family of Villa La Fonte, including researchers from higher and lower cohorts, visiting researchers, staff and professors. I cannot imagine a better environment to pursue a PhD.

Finally, my most special thank goes to my parents and my sister, for their efforts, which allowed me to pursue a PhD, and their unconditional love and support. I also want to thank my grandparents, for their immeasurable pride in my accomplishments and for always being curious about my research. I want to thank all my dear friends in Eibar, especially Ainhoa, Ane, Maria, Miren, Nerea and Onintza. Knowing that I always had you on the other side made this journey much easier. I also want to thank Ainhoa Osés for supporting me along the way. Lastly but certainly not least, I want to express my deepest gratitude to Konuray, the person who has been my rock through the thrilling peaks and daunting valleys of this journey. It is truly a great fortune to joyfully celebrate the smallest achievements with you, and to feel your love and belief in me during moments of self-doubt. I feel very lucky to be able to grow together on this journey. I dedicate this thesis to all of you.

Contents

1

Determinants of Migration Choices: The Role of Beliefs about Pecuniary and Nonpecuniary Outcomes

Abstract This paper studies how expected pecuniary and nonpecuniary factors influence migration decisions of tertiary educated young adults in advanced economies. Using a rich dataset on subjective expectations collected during the decision-making process, I estimate a life-cycle model of migration choice that takes expected migration duration into account. Crucially, the collected data allow me to separate preferences from beliefs and to distinguish between pecuniary and nonpecuniary factors. Regarding pecuniary factors, I find that migration decisions are more sensitive to earnings, followed by the prospects of full-time employment and a better match between studies and job. Although there is sorting on expected labor market outcomes, my results show that the set of nonpecuniary factors, such as being close to family and quality of social life, play a larger role in choosing whether to migrate. Finally, counterfactual exercises reveal that a human capital acquisition strategy has a limited effect on short-term migration plans, which are primarily driven by nonpecuniary factors.

1.1 Introduction

The distribution of economic activity is not uniform across locations (e.g., Moretti, 2013). Consequently, many individuals sort into the most productive areas, with higher wages and wage growth. Many of these migrations are only temporary and return migrants enjoy a wage premium to accumulated experience upon return (e.g., De la Roca and Puga, 2017). However, while many people move permanently or temporarily, most stay in their birthplace. The limited geographic mobility, along with rising regional inequalities in advanced economies, and the growing evidence on the crucial role of birthplace on lifetime economic outcomes (Chetty and Hendren, 2018) have prompted a renewed policy interest to address differences in opportunities by birthplace. These policy debates require to understand why some people migrate while others stay behind. This paper sheds light on this issue by estimating a life-cycle model of migration choice that uses rich data on subjective expectations about expected pecuniary and nonpecuniary factors.

Investigating the determinants of migration decisions using observed migration choices and realized outcomes (e.g., realized earnings) is challenging for at least four reasons. First, the typical dataset does not contain information about nonpecuniary outcomes, which have been long-acknowledged to be central to migration decisions (Sjaastad, 1962). Second, while expected net gains from migration are location-specific, the destinations that potential migrants consider are unknown to the researcher. Third, potential migrants' information sets are also unknown to the researcher. This lack of data makes the construction of beliefs for the non-chosen alternatives complex, since the well-documented selective sorting into migration implies that the formation of expectations about the returns to migration alternatives is arguably heterogeneous across groups. Fourth, it is the returns *perceived* by individuals that will influence actual migration decisions. Even if migrating were random, using realized outcomes to understand migration decisions requires the researcher to assume a mapping between realized outcomes and beliefs about them. This mapping is complex since the returns to migration that potential migrants anticipate are often biased (e.g., McKenzie et al., 2013). Overall, making inference on the decision-making process based on choice data and maintained (nonverifiable) assumptions on expectations is problematic since observed choices might be consistent with several combinations of expectations and preferences (Manski, 2004).

In this paper, I circumvent the identification problem by collecting data on individuals' subjective expectations about pecuniary and nonpecuniary outcomes at different future points in time under *each* migration alternative. Beliefs are fully heterogeneous, can be biased, and we remain agnostic about agents' information sets. Individuals are forward-looking and counterfactual alternatives include not migrating, migrating short-term (return migration) and migrating long-term, where migrating is defined as moving out of one's region of birth. These alternatives are mutually exclusive and constitute the complete choice set. For the migrating options, we elicit expected outcomes at each individual's most likely migration destination, i.e., the location where she would move to *if* she was to migrate, which can be another region within the country of birth or another country.

Pecuniary outcomes include employment status, wage conditional on employment status and the match quality between the job and the academic degree.[1] Nonpecuniary factors include enjoying being close to family, partner and friends, and the quality of social life. Crucially, all expected outcomes are elicited *before* the choice is made (when all individuals are still living in their region of birth) and they capture the uncertainty over the future stream of outcomes at the time of decision-making. Additionally, the survey collected individuals' stated choice probability of each migration alternative.

The survey design allows us, first, to directly estimate how agents -independent of selection into migration- expect different migration choices to affect their career profiles and social life outcomes over the life cycle. For example, we can estimate whether potential migrants anticipate a wage premium upon return. Second, by combining the expectations under the counterfactual scenarios with the expected choices into a discrete-choice model, we can estimate the extent to which the anticipated returns influence the decisions without making strong assumptions on expectations.

I collected data on 609 individuals born in Andalusia, Spain's most populous region with one of the highest unemployment rates in Europe. As is generally the case for other countries in Europe and the US, regions with higher unemployment rates in Spain also have systematically lower average wages and lower shares of high-skill jobs. I focus on individuals with tertiary education in school-to-work transition, for two reasons. First, this is the most mobile group of people -individuals are most mobile early in their careers and the low educated are much

[1] I use the terms pecuniary outcomes, labor market outcomes and career-related outcomes interchangeably.

less mobile, especially at the beginning of their careers (Amior, 2019). As I will argue later, focusing on the most mobile group of people in an area with few labor market opportunities makes it plausible to assume that all survey participants -including those who will choose to stay- have been actively thinking about migration decisions, which is implicitly assumed in the employed survey methodology. Second, this is a relevant population to focus on because despite most of the highly educated staying, lagging-behind regions worry about the negative consequences of disproportionately losing the relatively scarce highly educated workforce permanently (i.e., brain-drain).[2] The setting is relevant beyond Andalusia or Spain, as the large regional economic disparities in Spain are comparable to those in France, Italy or the US. The geography of Andalusia, surrounded by other regions with limited career prospects, implies that migrants choose long-distance moves -as opposed to daily commutes to bordering regions- making it an interesting setting to analyze migration.

The collected subjective belief data paint a sensible picture. 73% of students think that they would have the highest earnings over the life-cycle if they were to migrate long-term, which is consistent with migrating to improve the persistently poor labor market conditions in their region of birth. The data also show that students on average anticipate an earnings premium after return. 10 years after graduation, they expect to earn 19% higher earnings *in their region of birth*, Andalusia, if they have accumulated some working experience abroad (short-term migration) than if they always lived in their region of birth (no-migration). The data also reveal that students anticipate nonpecuniary outcomes to be affected by their migration choices. 90% of students expect a higher quality of social life at home than abroad, and they do not expect the gap to close as they accumulate years of life in their migration destinations.

I then combine subjective choice probabilities and subjective beliefs into a single coherent life-cycle model. The model includes expected earnings, expected study-job match prospects, expected enjoyment from being close to loved ones and expected enjoyment of quality of social life.[3] I find that all outcomes are statistically significant determinants of migration choices. In order to interpret their economic significance and compare results to other studies, I use the

[2]For example, brain drain may affect political polarization and social unrest in left-behind regions (Rodríguez-Pose, 2018; **?**) .

[3]Expected earnings are calculated by averaging earnings conditional on employment status with employment status probabilities. Throughout the paper, expected earnings refer to these weighted earnings.

estimates of the model parameters to calculate elasticities of choice with regard to pecuniary outcomes and willingness-to-pay estimates for nonpecuniary outcomes.

I estimate an average elasticity of choice with respect to earnings equal to 0.80, which is higher but similar in magnitude to other recent studies that analyze migration in Spain using aggregate data and alternative identification strategies and populations (see Melguizo and Royuela (2020); Clemente et al. (2016)).[4] The elasticity of choice to changes in full-time employment probabilities is 0.68, which is closely followed by the elasticity of choice with respect to having good study-job match prospects. Choice responses to increases in part-time employment probabilities are much lower, consistent with young adults at the start of their professional careers seeking to work full-time. On the other hand, I estimate that students have a total willingness-to-pay equal to 74% of their life-cycle expected earnings (12,185€ annually) to increase each of the nonpecuniary factors from their expected levels in the long-term migration alternative to their expected levels in the no-migration alternative. This number is considerably lower than the moving costs estimated in other studies, which exceed 100% of income (e.g., Kennan and Walker (2011), Ransom (2022)).[5]

I then use the model parameter estimates to perform a series of counterfactual exercises to asses the role of pecuniary versus nonpecuniary outcomes on expected migration choices. The first counterfactual equalizes students' beliefs about labor market outcomes across the three migration alternatives. That is, it assumes that students believe their migration choices will not affect their professional careers. The second counterfactual equalizes students' beliefs about nonpecuniary factors across migration alternatives, assuming they expect nonpecuniary factors to be unaffected by their migration choices. Comparing the changes in expected choices driven by each counterfactual scenario provides a meaningful metric to assess the role of pecuniary versus nonpecuniary factors on expected migration choices. Results show that young adults are more responsive to nonpecuniary factors than to labor market outcomes: for example, if

[4]The result is also comparable to the elasticity of the choice probabilities to changes in earnings found in other migration contexts, using other methodologies (e.g., Dahl and Sorenson (2010) find elasticities in the range of 0.5-1% for Danish engineers, and Bertoli et al. (2013) equal to 0.95% for migration choices from Ecuador to Spain).

[5]These studies estimate dynamic choice models and identify moving costs by assigning a distinct status to each person's birthplace. The discrepancy is likely due to a number of reasons. First, I am able to better identify the costs. Second, I estimate costs for individuals' chosen destinations relative to staying. Instead in their models, the moving cost represents the cost faced by the average individual if they were forced to move to an arbitrary location.

students believed that migrating did not affect their labor market outcomes, they would be 9 p.p. (25%) *more* likely to plan to stay. Instead, believing that they would enjoy the same nonpecuniary factors across alternatives would make them 14 p.p. (41%) *less* likely to plan to stay in their region of birth.

Finally, I do two counterfactual exercises to understand the drivers of planned short-term migration. Young adults in the sample expect short-term migration to be 40% more likely than long-term migration, a pattern observed in realized migrations of young adults from Andalusia to other Spanish regions. The counterfactuals manipulate beliefs in the periods in which in the short-term migration alternative students are back in their region of birth, after having migrated. The goal of the counterfactuals is to quantify the extent to which, given their preferences, young adults' plan to migrate short-term is motivated by (i) an anticipation of career benefits after return (i.e., short-term migration as a human capital acquisition strategy); (ii) low expected nonpecuniary factors under long-term migration. Results show that while both mechanisms exist, the second one plays a major role. I find that the choice of short-term migration would drop by 2 p.p. (4%) if individuals expected no career benefits after return. Instead, the choice of short-term migration would drop by 7 p.p. (17%) if expected nonpecuniary factors after return were equal to those expected in the long-term migration. This suggests that individuals' choice to return is more sensitive to nonpecuniary conditions at the migration destination (e.g., whether they make friends or find a partner at the destination) than to labor market conditions at home.

This is the first study that uses subjective expectations data to understand migration decisions under uncertainty. Previous migration studies have used subjective expectations data to asses the accuracy of individuals' expectations about actual realizations in the population, because systematic biases in beliefs can call for policy (information) interventions. These papers focus on migration from developing countries either using regular (e.g., McKenzie et al. (2013)) or irregular pathways (e.g., Bah and Batista (2020)), where information is scarce and particularly valuable.[6] This paper instead uses subjective expectations data to shed light on the determinants of migration choices, which has been traditionally answered using choice data.

[6] Other migration studies that use expectations data include Gibson and McKenzie (2011a) and Adda et al. (2022)

The paper contributes to and builds on three strands of the literature. First, it belongs to the long tradition of work seeking to understand whether expected labor market outcomes influence migration choices (e.g., Tunali (2000), Dahl (2002), Kennan and Walker (2011), Grogger and Hanson (2011), Gibson and McKenzie (2011a), Bertoli et al. (2013)). This research has used choice data and has mostly studied the role of expected earnings. I complement this research by introducing a new methodology to the migration literature, which uses individual expectations under counterfactual scenarios. This approach allows me to study the role of a broader set of career-related outcomes on young adults' migration choices, other than earnings. Moreover, it allows me to circumvent the identification problem concerning the separation of preferences and beliefs, present in studies that use choice data.

The paper also builds on the more recent literature that unpacks the black box of migration costs by measuring the role of nonpecuniary factors on migration choices. Dahl and Sorenson (2010), Huttunen et al. (2018) and Büchel et al. (2020) use observational data to understand individuals' preference to move close to family, friends or to places where their broader networks are. Using stated preference approaches, Koşar et al. (2021) focus on measuring the preferences for different location characteristics (e.g., crime rate) and moving costs, and Gong et al. (2022) characterize the total value of nonpecuniary benefits. I complement this literature by quantifying the importance of nonpecuniary factors by incorporating expectations about these factors as well as expectations about labor market outcomes directly into the choice model. My approach allows one to learn about individuals expectations about nonpecuniary factors, which is interesting per se, and takes individuals' expected migration duration into account.

Finally, this paper adds to the growing literature that uses subjective-expectations data to understand decision-making under uncertainty. The methodology that I employ has mostly been used to study educational choices (Boneva et al., 2022; Wiswall and Zafar, 2015, 2021) or occupational choices (Arcidiacono et al., 2020)[7]. It rests on the implicit assumption that the stated choices reported in the hypothetical scenarios are reflective of what respondents

[7]Other studies on educational choices (Arcidiacono et al., 2012; Attanasio and Kaufmann, 2014, 2017; Delavande and Zafar, 2019; Giustinelli, 2016; Kaufmann, 2014; Stinebrickner and Stinebrickner, 2014; Zafar, 2013) and health choices (Delavande, 2008) also elicit beliefs in counterfactual scenarios, but elicit only the alternative that individuals are most likely to choose or a ranking of them. This approach cannot capture individuals' uncertainty at the time of the survey (Blass et al., 2010), which is important in my setting, as revealed by the results.

would do in actual scenarios. There is growing evidence that the stated approach yields meaningful responses when the counterfactual scenarios presented to respondents are realistic and relevant for them (Wiswall and Zafar, 2021). Given that the survey is carried out at the time of making migration decisions -when they are about to graduate in a region with high migration prevalence- I argue that this is the case in this study. In this regard, students' expected migration choices are consistent with actual self-selection patterns observed for migrants in Spain: being younger, male, from higher socioeconomic status and having higher grades are all statistically and positively related to students' expected probability of migrating (González-Leonardo et al., 2022). I complement this literature by studying a new and relevant decision context, migration choices of young adults.

The rest of the paper is organized as follows. Section 1.2 outlines the model. Section 3.3 explains how I collected the data and section 1.4 describes the collected beliefs. Section 1.5 presents the life-cycle model's results and results from counterfactual exercises. Section 3.8 concludes.

1.2 Conceptual Framework

This section develops a simple model of migration choice at the beginning of individuals' labor market careers. The model's flexibility is based on the data I collect, which is described in detail in Section 1.4

1.2.1 Migration alternatives

Upon labor market entry, individual i at time t_0 considers the different migration alternatives that she could follow from t_0 to T, where T is equal to 10 years after finishing the bachelor's degree. The complete migration choice set is summarized in the following three alternatives:

No-migration, $m = 1$*: Always work in her region of birth.*

Short-term migration, $m = 2$*: Work outside for some time but return to her region of birth to work by period T.*

Long-term migration, $m = 3$*: Work outside for some time and NOT return to her region of birth to work before period T.*

After comparing how each migration alternative is expected to affect several relevant pecuniary and nonpecuniary factors from t_0 to T, individual i at time t_0 intends to follow the migration alternative that maximizes her expected utility.

Agents weight the trade-offs of the different alternatives at their most likely migration destination, *"the destination where they think they would migrate if they were to migrate"*. This destination can be anywhere outside their region of birth i.e., another region within their country of birth or another country. In the survey, those individuals who chose another region were further asked about which province and those who chose another country about which continent *[Europe/North America/South America/Africa/Asia/Oceania]*. The purpose of asking about the potential migration destination was to help participants more consistently envision their future experience (as in Boneva et al., 2022). During the survey, individuals were emphasized that they should imagine living in their reported migration locations.

By asking about one single migration destination the framework effectively assumes that individuals choose the same destination if they migrate short-term and long-term and that at the time of choosing whether to migrate or stay they do not *plan* onward migration i.e., moving from one migration destination to a new one, different from their region of birth. While this might be a strong assumption, the reason for limiting the choice to a single location is twofold. First, to reduce the risk of experimenter demand, as asking for more than one potential destination could make survey participants think about destinations they had never thought about before. Second, to reduce the cognitive load required to elicit expected outcomes in different locations.

1.2.2 Timing of belief elicitation

Beliefs are elicited at time τ when individual i is still a student close to finishing her bachelor's degree in her region of birth.

There are several factors that need to be considered when choosing the timing for eliciting beliefs. First, in order to minimize the risk of cognitive dissonance or ex-post rationalization (Festinger, 1957), beliefs should be elicited *before* choice-specific investments are made. In our context, this requires surveying individuals when they are still living in their region of birth. Second, one needs to take into account that migrating to work and migrating to undertake education are often interrelated (Dustmann and Glitz, 2011). As a result, there is a trade-off between measurement error and sample selection bias. On the one hand, surveying future college-educated individuals at the time in which they are finishing high-school in their region of birth would minimize sample selection bias, as individuals do not sort into high-school based on their propensity and motivation to migrate -they typically stay at their family home. At this stage, however, most students may have little idea of what migration alternative they want to pursue in their professional careers and may not have thought about the likelihood of the various outcomes conditional on migration alternatives (measurement error). On the other hand, surveying individuals when they are finishing the very last stage of their education (e.g., master's degrees) in their region of birth would minimize measurement error at the risk of incurring in sample selection bias by missing the early movers with highest propensity to migrate.

Therefore, in order to minimize the above-mentioned biases, I survey individuals shortly before finishing the bachelor's degree in their region of birth. Note that the fraction of students that pursue the bachelor's degree outside their region of birth is as low as 8% in our setting (FundaciónBBVA, 2018). Then, the survey collected data on students' expected highest level of education *[bachelor's degree/other type of further studies/master's degree/Ph.D.]* and the location *[region of birth/other region within country of birth/other country]* where they planned to pursue such studies.

Eliciting the probability of following each counterfactual alternative (instead of stating the most likely alternative) allows us to capture the fact that some uncertainty will be resolved on the value of the alternatives between the time in which beliefs are elicited at time τ and agents choose their intended migration alternative at time t_0. The subjective expected utilities of migration alternative m at times τ and t_0 are linked through the following relationship:

$$\mathbb{E}[U_{im}|\mathcal{I}_{it_0}] = \mathbb{E}[U_{im}|\mathcal{I}_{i\tau}] + \xi_{im\tau} \tag{1.1}$$

where $\mathbb{E}[.|]$ is the subjective-expectation operator, U_{im} is individual i's utility from migration alternative m and $I_{i\tau}$ and I_{it_0} are individual i's information sets at time τ and t_0 respectively. We denote as $\xi_{im\tau}$ the uncertainty to be resolved for migration alternative m when beliefs are captured at time τ (which Blass et al., 2010, refer to as "resolvable uncertainty"). The $\xi_{im\tau}$ term represents a preference shock that is assumed to be realized after students report their likelihood of choosing each migration alternative and before they actually make the migration choice. It reflects any unanticipated change in the utility of a migration alternative that occurs between τ and t_0 (e.g., finding a partner at birthplace).

Note that the subjective expected utility from migration alternative m elicited at time τ, $\mathbb{E}[U_{im}|\mathcal{I}_{i\tau}]$, accounts for *anticipated* utility changes resulting from actions that students plan to undertake between τ and t_0. The most obvious of these actions is further study plans after finishing their bachelor's degree. Individuals were instructed to consider in their response that they followed their reported study plans and therefore, the expected utility from migration alternative m at time τ, $\mathbb{E}[U_{im}|\mathcal{I}_{i\tau}]$, incorporates how students expect their study choices prior to t_0 to affect the utility of each migration alternative. Because future study and migration choices are sequential in time, each student evaluates the trade-offs from the three counterfactual migration alternatives under the same study plan.

1.2.3 Expected utility by migration alternative

At time τ, student i possesses a distribution of beliefs $G_{i,\tau}(x|m.t)$ about the probability of the vector of future outcomes $x \in X$ occurring in all future periods $t \geq t_{i0}$ *if* she were to choose migration alternative m. We allow the start period of the labor mobility alternative t_0 to vary across individuals with their expected maximum level of education. Student i's subjective expected utility from migration alternative m at time τ is given by

$$\mathbb{E}[U_{im}|\mathcal{I}_{i\tau}] = \sum_{t=t_{i0}}^{T} \beta^{t+g_i} \int_{x \in X} u(x) dG_{i\tau}(x|m,t) \tag{1.2}$$

where $t_{i0} \in \{0,1,2,3\}$. We set $t_{i0} = 0$ if student i does not plan to pursue further studies after finishing the bachelor's degree and $t_{i0} = 1$, $t_{i0} = 2$, $t_{i0} = 3$ if they plan to pursue other type of studies, a master's degree and a PhD respectively. This specification reflects that the migration alternative of students who plan to pursue further studies is shorter in length and starts farther in time relative to students who plan to enter the labor market right after finishing their bachelor's degree. $g_i = \{0,1,2,3\}$ is the student i's years until graduation, with $g_i = 3$ if the student is a freshman, $g_i = 2$ (sophomore), $g_i = 3$ (junior) and $g_i = 0$ (senior). $\beta \in (0,1)$ is the discount rate. $u(x)$ is the migration alternative's utility function that provides the mapping from the finite vector of outcomes x to utility. A key feature of the model is that when choosing the intended migration alternative m at time t_0, the student faces uncertainty about the labor market outcomes and nonpecuniary factors over the life-cycle. For tractability, we assume that the utility function is additively separable in pecuniary and nonpecuniary attributes.

$$\begin{aligned}\mathbb{E}[U_{im}|\mathcal{I}_{\tau}] = &\sum_{t=t_{i0}}^{T} \beta^{t+g_i} \Big[\phi_1 \sum_{l=FT,PT} P_{imt}(L=l) * \mathbb{E}(w_{q,imt}|L=l) + \phi_2 P_{imt}(\text{Study-Job Match})\Big] \\ &+ \sum_{t=t_{i0}}^{T} \beta^{t+g_i} \Big[\phi_3 P_{imt}(\text{Enjoy social life}) + \phi_4 P_{imt}(\text{Enjoy being close})\Big] + \gamma_m\end{aligned} \tag{1.3}$$

Pecuniary outcomes (labor market outcomes): where $\sum_{l=FT,PT} P_{imt}(L=l) * \mathbb{E}(w_{q,imt}|L=l)$ are expected earnings, that is, expected earnings conditional on employment status averaged by expected employment status probabilities. $P_{imt}(L=l)$, where $l = \{FT, PT\}$, is student i's expected probability of working full-time and part-time in migration alternative m at time t and $\mathbb{E}(w_{q,imt}|L=l)$ is students i's expected minimum (maximum) yearly gross earnings in migration alternative m at time t conditional on employment status l, where $q \in \{min, max\}$. Following Wiswall and Zafar (2015), the survey asked respondents for their potential earnings if they were working full-time and asked their beliefs about the probability of working full-time, part-time, being out of the labor market, and being unemployed as a separate question. Collecting expected wages conditional on employment status on the one hand and employment status probabilities on the other allows me to circumvent the standard endogenous selection into employment issue where job characteristics are only observed for individuals who work. The utility model specification assumes that students value expected earnings, which can operate through the earnings offered if employed or the employment status probabilities. Some students, for example, could think that conditional on getting a job, wage differences would not be big between their chosen migration destination and home, but that the chances of getting a job are much lower in their region of birth. Because the survey asked about full-time equivalent wages only, I assume that students' expected wages if working part-time are half of those if working full-time. I also assume that students have no monetary compensation when unemployed or out of the labor market.[8] I specify two different utility functions, one in which students care about the potential wages' lower bounds, w_{min}, and another one in which migration decisions are driven by the highest earnings that they could think of having, w_{max}. $P_{imt}(\text{Study-Job Match})$ is student i's expected probability of having a job that is directly related to her bachelor's degree at time t in migration alternative m, conditional on being employed. Expected earnings and study-job match make up the pecuniary or career-related outcomes that students value in a migration alternative, and ϕ_1 and ϕ_2 are the utility values from the discounted expected lifetime values of these outcomes.

Nonpecuniary outcomes: The set of nonpecuniary outcomes is formed of $P_{imt}(\text{Enjoy social life})$ and $P_{imt}(\text{Enjoy being close})$, which stand for student i's expected probabilities of enjoying the quality of their social life and enjoying being close to family members, partner and friends at time t under migration alternative m. Absent direct measures on respondents' beliefs about nonpecuniary outcomes, researchers typically introduce a "home-bias" in the utility specification, understood as a utility cost of living away from one's birthplace.[9] With direct individual measures, I let each student in the sample express whether they perceive a decrease or an increase in the probability of enjoying the quality of their social life or in the probability of enjoying being close to loved ones when moving abroad. This flexible specification allows individuals to perceive, for example, no trade-off between gains

[8]This pulls the probability of being unemployed and out of the labor market together. The probability of the latter is tiny in all the three alternatives, as shown in section 1.4.

[9]See for example, Kennan and Walker (2011); Ransom (2022); Zerecero (2022)

in pecuniary outcomes and quality of social life if they were to migrate instead of staying in their region of birth. Similarly, the specification accounts for the fact that some students can enjoy being close to their loved ones not only if they stay in their region of birth but also if they choose to move to a destination where their loved ones are already living, which is plausible in places with high migration prevalence if new migrants follow previous movers. ϕ_3 and ϕ_4 are the utility values from the discounted expected lifetime value of these two nonpecuniary factors. γ_m captures choice-specific unobservable factors that affect lifetime utility. Our goal is to estimate the parameter vector $\Theta = \{\{\phi_j\}_{j=1}^4, \gamma_m\}$ up to scale.

1.3 Data

The data is from two original online surveys administered to last year undergraduate students in the region of Andalusia (Spain) over 3 weeks during June 2020.

1.3.1 Institutional Setting

Spain is comprised of 17 regions and Andalusia is the southernmost, most populous (over 8 million inhabitants) and second largest (over 87,000 km^2) region of the country. It is also the region with persistently highest youth unemployment rate, lowest wages and overall poor labor market conditions. As shown in Figure 1.1, the lack of labor market opportunities in a region strongly correlates with its population's willingness to reallocate for work.

Given that its bordering other southern regions of Spain also lag behind the national average, migrations out of the region typically involve long distance moves -as opposed to commutes-.[10] 52% of survey participants chose Madrid (capital city of Spain, about 500km away from Andalusia and 5 hours by car) as the destination where they would migrate if they were to migrate, 28% chose a country in Europe and 6% Barcelona (second largest city of Spain, about 1000 km away from Andalusia and 10 hours by car) -all other destinations were chosen by 2% or less of the sample-.

[10] The south of Spain has been historically poorer, with persistently lower income and employment rates. To this day, per person net income in the south is about 60% of that in the center-north, and unemployment rates are around double.

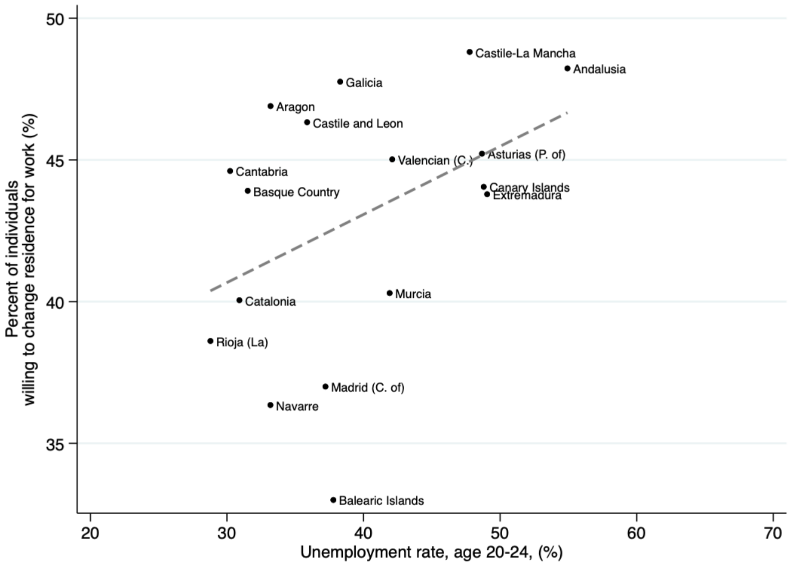


Figure 1.1 Fraction of individuals willing to move if they are provided with a job by unemployment rate of residency

Migration patterns from the region are in line with those of other developed economies. The hazard of migrating is highest for individuals between 23-28 years old and individuals with tertiary education are twice as likely to migrate as their lower educated counterparts (Table A.1) (e.g., Amior, 2019, for the U.S.). While some of these young and highly educated individuals migrate from their region of birth permanently, temporary migration is 150% more likely than long-term migration in the first decade of career trajectories: 10 years after migrating, 50% of the migrant cohort has returned back to their region of birth. The high prevalence of temporary migration is a feature of other developed countries too (e.g., Bartolucci et al., 2018a; Dustmann and Görlach, 2016).

The fact that 85% of survey participants claim to have a close friend living outside Andalusia and that, among those who have older siblings who are working, 18% and 15% report that these are working in another Spanish region and in another country respectively shows that participants are exposed to migration. Indeed, 35% of participants answered that it was easy to imagine the hypothetical scenarios either because it was a daily topic of their lives (86%) or because they had been getting informed (14%). 63% responded that it was difficult but that the hypothetical scenarios involved questions that they had been thinking about beforehand and that it was therefore a relatable topic. Only 2% of survey participants answered that they hypothetical scenarios were difficult to answer and that they hadn't been actively thinking about such topic.

I take this as evidence that the counterfactual scenarios presented to respondents -which involve migration choices to their chosen migration scenarios- are realistic and relevant for young and highly educated individuals living in advanced economies with few labor market opportunities. There is growing evidence (e.g., Delavande and Zafar, 2019; Wiswall and Zafar, 2015) that the stated approach yields meaningful responses when this is the case. As I clearly show in section 1.4, the information that respondents provide is sensible and meaningful.

1.3.2 Survey Administration

The two surveys were designed using Qualtrics Survey software and students received the online survey links via email two days apart. The surveys were administered at two large public universities that rank highest among Andalusian universities according to the QS World University Ranking i.e., the University of Seville (US) and the University of Granada (UGR). Map A.2 shows the location of the two universities. The study was restricted to the schools of Social Science and Law, Engineering and Architecture, and Natural Science within these universities (table A.4 in the Appendix reports the list of degrees that participated in the survey). I excluded the fields of Health Science and Arts and Humanities because many of these students end up working in the public sector (e.g., as doctors or teachers) and their migration choices are often determined by the results from competitive exams to become public servants.

The target population were last year bachelor's degree students in the schools described above, which included 5,296 students across the schools. Students received the email with the link to the first survey directly from the schools through their official communication channel. The survey link was closed 12 hours later, once the number of participants that our budget permitted to compensate was reached. This corresponded to a response rate of 18%.[11] The first survey collected students' individual and family characteristics, as well as their contact information (phone number and email address). The survey took approximately 7 minutes (median) to complete. At the end, it included a luck game with probabilistic payoffs

[11]This response rate for the survey is broadly comparable to that of other surveys conducted on similar populations - for instance, the response rate for Cortés et al. (2022)'s survey of Boston University's Questrom School of Business was 20% while the response in the Global COVID-19 Student Survey by Jaeger et al. (2021) was around 10% to 12% across the 28 universities that participated.

that encouraged participation. The payments were transferred within the next 72 hours of completion using a popular online payment platform in Spain (Bizum).[12]

Students were told that finishing the first survey was necessary but not sufficient to be invited to the second survey, without revealing the condition for being invited. Only students who reported being born in Andalusia were invited to participate in the second survey (688 students). As mentioned in section 1.2.2, focusing on individuals studying in their region of birth minimizes the risk of ex-post rationalization, as beliefs are elicited before migration costs are incurred. Students were sent the link to the second survey to the email they had provided in the first survey and were also notified about the arrival of the email via sms. 90% of the invited students completed the follow-up survey.

The second survey took approximately 25 (median) minutes to complete, and students were compensated with 6€ for successfully completing it. Students were given one week to start the survey before the link became inactive and were told to complete the survey in one sitting. Halfway through the week, those who had not completed the survey were sent a reminder. The second survey collected all remaining data: study plans, expected migration choices, and beliefs under the counterfactual migration alternatives. As with the first survey, payments were done via Bizum within 72h of completion. Section A.0.1 in the Appendix describes the survey administration in more detail.

1.3.3 Sample Characteristics

The final sample consists of 609 individuals. Table 2.1 shows the sample's descriptive statistics. It also shows how students in the sample relate, in terms of educational characteristics, to the population of students studying the same degrees in the same departments and, in terms of family background characteristics, to the population of Andalusian university students. Results in Table 2.1 show that the fraction of females, the mean age at degree completion, and the average GPA in my sample are very similar to the ones in the population of students enrolled in the last year of equivalent degrees, i.e., 46% vs. 48% of females, 23 vs. 24 years old, and 6.82

[12]Bizum is a highly popular instant and commission free phone to phone payment system that was launched by Spanish banks in 2016. It is used by friends, colleagues or family members to make small instant payments. All what is needed to receive money is a bank account, the Bizum app - which was offered by 96% of all Spanish banks in 2020- and a phone number. Knowing the phone number of the recipient is enough for the sender to make the transfer.

vs. 6.9 GPA in my sample relative to the administrative data respectively. The administrative data come from the Spanish Ministry of Universities and are publicly available.[13] The survey

Table 1.1 Sample characteristics and comparison to other data sources

	My sample		University and LFS data
	Mean (s.d.)	N	Mean
Last Year (%)	81	609	
Field of Study (%)			
- Natural Science	8	609	
- Social Science & Law	63	609	
- Engineering & Architecture	29	609	
Female (%)	46	609	48
Age at survey completion	23.09	609	
	(2.23)		
Expected age at end of bachelor's degree	23.30	609	24
	(2.27)		
GPA, 0-10 scale	6.82	534	6.9**
	(0.84)		
High SES (%)	49	601	46
Parents' employment status (%)			
- Both parents working	50	609	38
- One parent working	38	609	46
- Both parents unemployed	2	609	1
- One parent unemployed	16	609	14
Perceived family wealth w.r.t. the average family in Andalusia (%)			
- Higher or much higher	30	600	
- Similar	56	600	
- Lower or much lower	14	600	

Note: Last year refers to the fraction of students enrolled in their last year of the bachelor's degree. Students are categorized as high socioeconomic status (SES) if at least one of their parents attended university, and low-ses otherwise.

also collected data on students' family background. In particular, about parents' highest level of education, employment status, and perceived family wealth relative to the average family in Andalusia. 49% of students in the sample have high socioeconomic status (defined as having at least one parent with a university degree), 50% have both of their parents working, 16%

[13]See link for individual characteristics this other link for grades by university and degree.

have one of their parents unemployed, and 30% think that their family wealth is higher than the wealth of the Andalusian average family. To assess how my sample's students' household characteristics relate to families who have children attending university in Andalusia, I use the Spanish Labor Force Survey (LFS), year 2019. I define a comparable family as one that (i) is living in Andalusia and (ii) has at least one child born in Andalusia and attending university in 2019. Overall, the results show that the family characteristics of students in the sample are similar to those of Andalusian families that have children attending university that year. For example, 46% of these students in the LFS and 49% of students in my survey have a high socioeconomic status.

1.4 Description of Beliefs

This section describes the subjective expectations data.

1.4.1 Expected migration choices

Table 1.2 reports the distribution of students' reported subjective choice probabilities for each migration alternative (Figure A.3 plots the kernel density distributions).

Table 1.2 Subjective choice probabilities, (%)

	No-migration	Short-term migration	Long-term migration
Mean	36.5	36.7	26.7
Sd	27.2	20	21.2
10th percentile	5	10	5
25th percentile	10	20	10
50th percentile	30	35	20
75th percentile	55	50	40
90th percentile	80	60	55

Results show that students on average view migrating temporarily as likely as staying in their birthplace, while following a long-term migration alternative is viewed as much less likely. Students at this stage are uncertain about their future migration choices, highlighting the benefit of using stated-probability choices as opposed to stated choices (Blass et al., 2010).

Results also show that the distribution of choice probabilities is different across alternatives: short-term migration follows a symmetric distribution, while the no-migration and long-term migration alternatives are right skewed.

1.4.2 Expectations about pecuniary outcomes

Next, I describe students' subjective expectations about labor market outcomes under each migration alternative. Due to time and respondent burden considerations, one cannot ask respondents to report their beliefs for every year of the migration alternative. Instead, the survey asked these beliefs for two future points in time -3 and 10 years after finishing the bachelor's degree. The survey instructed students to assume that by $t = 3$, they had finished all their studies and were already living at their chosen migration destination (in the short-term and long-term migration alternatives). Expectations about all labor market outcomes were collected in tables such as the one below

Table 1.3

			$t = 3$	$t = 10$
$m = 1$: Region of birth	-	Region of birth	—	—
$m = 2$: [Migration destination]	-	Region of birth	—	—
$m = 3$: [Migration destination]	-	[Migration destination]	—	—

This table lets students easily compare expected outcomes at each period in time across alternatives taking the *complete location sequences* that comprise each alternative into account. By eliciting beliefs in this manner, we can construct measures of interest that are typically not observed. For example, we can directly construct, with no parametric assumptions and taking the beliefs of *all* individuals into account, regardless of their sorting plans, the expected wage premium after return. This is the wage difference that students expect 10 years after graduation when they are living in Andalusia, their region of birth, after having accumulated some working experience by working in Madrid relative to expected earnings if they have always worked in Andalusia. This is therefore a direct measure of whether young adults view temporary migration as a human capital acquisition strategy. The individual-specific expected wage premium after return is computed as:

$$\delta_{i,m_2,t_{10}} = \frac{\mathbb{E}(w_i|m=2,t=10) - \mathbb{E}(w_i|m=1,t=10)}{\mathbb{E}(w_i|m=1,t=10)}$$

The remaining expected wage premia are computed in a similar manner. $\delta_{i,m_3,t_{10}}$ is the expected percent wage difference 10 years after graduation, if the student is living in Madrid and has been living there for at least 8 years, as opposed to if she is living in Andalusia and has always lived there. δ_{i,m_2,t_3} and δ_{i,m_3,t_3} are the short-term premium i.e., 3 years after graduation, of short-term (temporary) and long-term migration, respectively, relative to no-migration. δ_{i,m_2,t_3} and δ_{i,m_3,t_3} will differ if students expect to have a different behavior when they arrive to their migration destination (e.g., make different job search effort, sort into different occupations) depending on whether they plan to return back home or stay in the migration destination for a long-term. [14]

This section describes students' beliefs about earnings, employment status and study-job match prospects 3 and 10 years after graduation by migration alternative (as in Table 1.3). Note that to elicit these beliefs, students were told to think about jobs that they thought they would be offered and that they would accept (as done in previous surveys that elicit this type of questions (Jensen, 2010; Wiswall and Zafar, 2015)). The section also reports expected premia for different outcomes and future points in time as described above. The distributions of these expected premia and the tables reporting the different moments of the distributions are reported in Appendix.

Expected monthly minimum and maximum gross earnings if working full-time: Students were asked about the expected minimum and maximum monthly gross earnings *if* working full-time, and they were asked about the probability that they would be working full-time as a separate question. This allows to elicit beliefs about all outcomes to every individual. Students were asked about monthly -rather than annual- earnings because individuals, especially those who do not have a work contract yet, are most used to referring to wages on a monthly basis in Spain. When I refer to annual wages, these are monthly wages multiplied by

[14]In practice, more than at arrival, the consequences of a different behavior would be more likely to emerge in the longer term, for which one would have to elicit beliefs for more than one future point in time at the migration destination. Adda et al. (2022) document that migrants who plan to stay longer at the destination invest more into skill acquisition than those who plan to return, which affects the career paths of the two groups at their migration destination.

12. Table 1.4 shows the moments of the distribution of expected monthly minimum and maximum gross earnings 3 and 10 years after graduation from the bachelor's degree by migration alternative.

Table 1.4 Sample distribution of expected monthly gross earnings, in €, if working full-time

		No-migration (1)		Short-term m. (2)		Long-term m. (3)	
		Min.	Max.	Min.	Max.	Min.	Max.
Panel A	3 years after						
	Mean	1,122	2,055	1,429	2,533	1,461	2,608
	First Quartile	900	1,300	1,000	1,500	1,000	1,500
	Median	1,000	1,600	1,200	2,000	1,300	2,000
	Third Quartile	1,200	2,300	1,700	3,000	1,800	3,000
	Standard Deviation	596	2,371	955	2,915	960	2,912
Panel B	10 years after						
	Mean	1,743	3,498	1,982	3,807	2,336	4,418
	First Quartile	1,200	2,000	1,300	2,000	1,500	2,500
	Median	1,500	2,500	1,700	3,000	2,000	3,100
	Third Quartile	2,000	3,500	2,300	4,000	2,500	5,000
	Standard Deviation	1,137	4,536	1,591	4,651	2,478	4,915

-

Panel A shows that three years after graduation, the mean of the expected minimum gross earnings if working full-time is 1,122€ if they stayed in their region of birth and 1,429€ and 1,461€ if they migrated to the place of their choice with the intention to return and to stay for a longer term, respectively. Students' expected maximum earnings follow the same pattern as minimum earnings. Its value is lowest for the no-migration alternative, 2,055€, and higher for the short-term and long-term migration alternatives (equal to 2,533€ and 2,608€ respectively). Students' expected earnings range (difference between minimum and maximum expected earnings) is smallest in their region of birth, and the low variance in the distribution of expected minimum earnings is remarkable. In particular, the median student believes that the minimum wage she would be working at in her region of birth is 1000€, close to the 900€ minimum wage set in Spain at the time of the survey. Figure A.4 shows the probability density functions of expected minimum earnings 3 and 10 years after finishing the bachelor's degree, and figure A.5 the kernel densities of the minimum earnings premia from choosing short-term and long-term migration alternatives relative to the no-migration alternative. This figure shows that the long-term migration premium is shifted to the right, both 3 and 10 years after finishing the bachelor's degree. That is, students think that at the beginning of their career, in the migration destination, they would earn higher full-time wages if they were to stay for the long-term instead of the short-term. The difference between the two, however, is not economically significant. The average expected minimum earnings premia across individuals are equal to 29% and 32% from short-term and long-term migration, respectively, and both are statistically significantly different from zero at 1% levels. That is, students believe that their earnings potential would have a sizeable increase 3 years after if they were to migrate both short-term and long-term. This results are reported in Panel A of table A.5.

Panel B in table 1.4 shows the same moments of the full-time earnings distributions, for 10 years after graduation. The mean of the expected minimum earnings if working full-time in their region of birth is 1,743€ and 1,982€ if they had never left the region and if they returned after having migrated temporarily respectively. This shows that students, on average, do anticipate a wage premium after return. The average expected premium, reported in table A.5, is 13%. The high spike around zero that shows the kernel density of this premium (figure A.5), however, reveals that many students expect no gains to having migrated (indeed, 40% of students expect exactly zero gains in full-time earnings and almost 10% expect negative returns, Figure A.6). Unsurprisingly, students expect their earnings potential to be highest 10 years after finishing the bachelor's degree if they were to migrate long-term, expecting their minimum earnings if working full-time to be, on average, 2,336€ a month. The expected average full-time earnings premium from long-term migration 10 years after is equal to 31%. This average premium is statistically significantly different (at 1% level) from the average premium that students expect from short-term migration 10 years after.

Expected employment status: The survey asked students about their expected probability of working full-time, working part-time, being unemployed and being out of the labor market 3 and 10 years after finishing the bachelor's degree if they were to follow each counterfactual migration alternative. Table 1.5 reports the mean, median and standard deviation of the responses to this question. Results show a rather pessimistic view on their employment opportunities. The average student believes that 3 years after finishing the bachelor's degree she has a 49% chance of working full-time and 28% and 17% chances of working part-time and being unemployed if she were to stay in her region of birth. Students expect higher chances of working full-time and lower chances of being unemployed if they were to migrate, while they expect their migration choices to have little effect on the probability of working part-time and being out of the labor market. Panels B, C and D of table A.5 report the employment status premia. The short-term and long-term migration premia of working full-time and being unemployed 3 years after show that on average students believe migrating short-term will increase their chances of working full-time by almost 47% and will decrease chances of being unemployed by 28% (both statistically significantly different from zero at 1% level). Results are similar for long-term migration. Over time, students expect their chances of being employed full-time to increase and the chances of working part-time and being unemployed to decrease in all migration alternatives. Students on average believe that the migration experience accumulated abroad will be of less help for getting a full-time job back at return in their home region. Finally, 10 years after finishing the bachelor's degree, students expect 13% higher chances of working full-time if they followed the long-term migration alternative than if they never migrated (statistically significantly different from zero at 1% level).

Table 1.5 Labor supply and study-job match beliefs 3 and 10 years after finishing the bachelor's degree by migration alternative

	Employment Status								Study-Job match	
	Work F.T.		Work P.T.		Unemployed		Out of lbr mkt			
	3 y.a.	10 y.a.	3 y.a.	10 y.a	3 y.a.	10 y.a	3 y.a.	10 y.a	3 y.a.	10 y.a
No-migration	0.49	0.70	0.28	0.18	0.17	0.08	0.06	0.04	0.55	0.74
	[0.50]	[0.70]	[0.30]	[0.15]	[0.15]	[0.05]	[0.00]	[0.00]	[0.50]	[0.80]
	(0.26)	(0.22)	(0.17)	(0.15)	(0.16)	(0.10)	(0.09)	(0.08)	(0.26)	(0.22)
Short-term m.	0.59	0.71	0.26	0.19	0.11	0.08	0.04	0.03	0.64	0.76
	[0.60]	[0.75]	[0.25]	[0.17]	[0.10]	[0.05]	[0.00]	[0.00]	[0.60]	[0.80]
	(0.26)	(0.23)	(0.18)	(0.16)	(0.11)	(0.11)	(0.08)	(0.07)	(0.24)	(0.21)
Long-term m.	0.59	0.74	0.26	0.18	0.10	0.06	0.04	0.02	0.65	0.79
	[0.60]	[0.80]	[0.25]	[0.14]	[0.10]	[0.03]	[0.00]	[0.00]	[0.70]	[0.85]
	(0.27)	(0.22)	(0.19)	(0.17)	(0.12)	(0.08)	(0.08)	(0.06)	(0.24)	(0.21)

Medians in brackets and standard deviations in parenthesis.

Expected study-job match: Young individuals in the sample have studies in the fields of Social Science and Law, Engineering and Architecture and Natural Science. Andalusia, on the other hand, is the Spanish region with the highest negative gap between high-skilled employment and total employment as a share of national employment OECD (2020). While this is largely the result of the level of education of its inhabitants, it is likely that due to the productive activity of the region many young skilled individuals seek to improve their study-job match prospects by migrating. The survey asks students about their expected probability of working in a job directly related to their bachelor's degree studies. The last two columns of table 1.5 report the responses to this question. The table shows that students, on average, expect higher chances of working in jobs directly related to their field of study if they were to migrate short-term, and specially long-term, than if they were to stay. 3 years finishing the bachelor's degree, they expect on average 41% and 46% higher chances of having a good study-job match in these two alternatives respectively. The difference, however, shrinks over time, as students expect their match prospects to improve over time, in particular, in their region of birth. As with potential earnings, students expect on average temporary migration to help them have better matched jobs back in their region of birth: on average, 9% premium, statistically significantly different from zero at 1% level. Thus, better expected study-job matches at return could in part explain the expected wage premium after return. 10 years after finishing the bachelor's degree, students on average expect 18% premium in the probability of having a job directly related to their bachelor's degree if they are living at their chosen migration destination as opposed to in their region of birth. This number is statistically significantly different from zero at 1% level. This measures are reported in Panel E of table A.5.

1.4.3 Expectations about nonpecuniary outcomes

In order to decrease the burden of questions and minimize fatigue, students were asked about nonpecuniary outcomes by location as opposed to migration alternative. In particular, students were asked about the probability of enjoying the quality of their social life in their region of birth and their chosen migration destination 3 and 10 years after finishing the bachelor's degree and about their expected probability of enjoying being close to family, partner, and friends 3 years after in their home region and their reported migration destination. This means that I assume that 3 years after finishing the bachelor's degree and living at their migration destination, students expect to enjoy the nonpecuniary outcomes the same whether they plan to return or stay long-term. Similarly, 10 years after finishing the bachelor's degree and living in their region of birth, they expect to enjoy the nonpecuniary outcomes the same regardless of whether they have migrated and returned or they have never migrated. Results show that students expect the quality of social life and enjoying being close to loved ones to be higher

Table 1.6 Beliefs about nonpecuniary outcomes

	3 years after		10 years after
	Social Life	Close	Social Life
Andalusia	0.80	0.85	0.67
	[0.80]	[0.90]	[0.70]
	(0.21)	(0.20)	(0.24)
Migration destination	0.59	0.34	0.50
	[0.60]	[0.30]	[0.50]
	(0.24)	(0.24)	(0.24)

Medians in brackets and standard deviations in parenthesis.

home than at their chosen migration destination. Note that while their expected probability of enjoying the quality of social life decreases over time, it decreases similarly at home and abroad. That is, students do not expect to adapt to their migration destination, in the sense that they do not expect the gap to close between home and abroad over time. The percent loss (negative premium) in social life enjoyment reported in table A.6 is on average 20% and 23%, 3 and 10 years after finishing the bachelor's degree, respectively (both statistically significantly different from zero at 1% level). The expected difference between home and abroad is higher for enjoying being close to family, partner and friends. The percent loss, reported in table A.6 is 56%. Note, however, that the expected probability -in levels- abroad is not zero (the mean is 34% and the median is 30%, Table 1.6). This suggests that students, to an extent, choose migration destinations where they have friends, partners, or family members.

1.5 Model Results

This Section presents the parameter estimates of the utility function of the model presented in Section 1.2, using the expectations data, interpolated over the life-cycle (see A.0.4 for a detailed explanation of these approximations).

1.5.1 Estimation of the preference parameters

We begin by assuming that the preference shocks described in equation (1.1) are perceived to be independent and identically distributed across individuals and migration alternatives following a standard type I extreme-value distribution. Then, student i's subjective probability at time τ of following migration alternative m is given by

$$\begin{aligned} p_{im\tau} &= P_i(\mathcal{E}[U_{im}|\mathcal{I}_\tau] + \xi_{im\tau} > \mathcal{E}[U_{in}|\mathcal{I}_\tau] + \xi_{in\tau}, (m,n) \in M_i, m \neq n) \\ &= \frac{exp(\mathcal{E}[U_{im}|\mathcal{I}_\tau])}{\sum_{n \in M_i} exp(\mathcal{E}[U_{in}|\mathcal{I}_\tau])} \end{aligned} \tag{1.4}$$

Taking the "no-migration" alternative, denoted as $m = 1$, as the reference alternative, we can re-write the log relative probability of choosing migration alternative m relative to migration alternative $m = 1$ as

$$\begin{aligned} ln\left(\frac{\tilde{p}_{im\tau}}{\tilde{p}_{i1\tau}}\right) = &\phi_1 \sum_{t=t_{i0}}^{T} \beta^{t+g_i}[\Delta \sum_{l=FT,PT} P_{imt}(L=l) * \mathbb{E}(w_{q,imt}|L=l)] \\ &+ \sum_{j=2}^{4} \phi_j \sum_{t=t_{i0}}^{T} \beta^{t+g_i}[\Delta P_{imt}(x=x_j)] + \gamma_m + \omega_{im} \end{aligned} \tag{1.5}$$

where Δ denotes the differencing operator taken with respect to the baseline migration alternative. $\sum_{l=FT,PT} P_{imt}(L=l) * \mathbb{E}(w_{q,imt}|L=l)$ are expected earnings, where $P_{imt}(L=l)$, $l = \{FT, PT\}$ is student i's expected probability of working full-time and part-time in migration alternative m at time t and $\mathbb{E}(w_{q,imt}|L=l)$ is students i's expected minimum (maximum) yearly gross earnings in migration alternative m at time t conditional on employment status, where $q \in \{min, max\}$. x_2, x_3, x_4 are the probability of having a job directly related to the bachelor's degree studies, the probability of enjoying the quality of social life, and the probability of enjoying being close to family, partner and friends respectively. γ_m for the no-migration alternative is normalized to zero. ω_{im} represents a measurement error, which reflects that the reports of migration alternative probabilities in our data, $\tilde{p}_{im}$, measure the "true" probabilities, p_{im}, with some error.

We now have a linear relationship between the known quantities in the data. We estimate the linear regression using ordinary least squares (OLS) and least absolute deviation (LAD) estimators. For estimation using OLS, I recode all reported extreme probabilities of 0 and 1 to

0.001 and 0.999, respectively.[15] Take into account that while the OLS estimator is sensitive to these roundings, the quantile estimator is not, and therefore, it is preferred Blass et al. (2010). Since we have two observations per respondent, standard errors are clustered at the individual level.

1.5.2 Model Estimates

Table 1.7 presents the LAD and OLS estimates of the utility specification in equation (1.3). All results assume that $\beta = 0.95$. Columns 1 and 3 show the results for the utility specification where students value expected minimum earnings, while columns 2 and 4 provide the counterpart results using maximum earnings.

Table 1.7 Estimates of Model Parameters

	(1) LAD	(2) LAD	(3) OLS	(4) OLS
ϕ_1x$10,000$: Minimum earnings	0.124***		0.132***	
	(0.0247)		(0.0268)	
ϕ_1x$10,000$: Maximum earnings		0.0697***		0.0590***
		(0.0150)		(0.0124)
ϕ_2: Prob. of good study-job match	0.204***	0.204***	0.311***	0.322***
	(0.0783)	(0.0765)	(0.0901)	(0.0889)
ϕ_3: Prob. of ejoying social life	0.352***	0.348***	0.386***	0.371***
	(0.0636)	(0.0634)	(0.0890)	(0.0890)
ϕ_4: Prob. of enjoying being close	0.135***	0.130***	0.157***	0.160***
	(0.0436)	(0.0459)	(0.0570)	(0.0565)
γ_{ST}	0.371***	0.401***	0.347***	0.336***
	(0.114)	(0.138)	(0.0937)	(0.0930)
Constant	0.0421	0.0442	0.132	0.220
	(0.214)	(0.238)	(0.198)	(0.193)
N	1134	1134	1134	1134
adj. R^2			0.157	0.159
pseudo R^2	0.081	0.081		

- Note: Parenthesis in OLS columns report robust standard errors clustered at the individual level. Parenthesis in LAD columns reported bootstrapped standard errors with 1000 replications clustered at the individual level. Minimum and maximum earnings are earnings conditional on employment status averaged by employment status probabilities.

In all four specifications, the expected minimum and maximum earnings coefficients are positive and statistically different from zero at the 1 percent level. The coefficients on minimum

[15] Note that in this context, the reports of extreme values, probabilities of exactly zero or one, reflect rounding, not censoring or truncation. In this context, there is little substantive difference between expressing a very low probability of following a migration alternative as 0.01 or zero. In any case, as shown in section 1.4.1, there are very few extreme cases in my data.

earnings are almost twice as large as those on maximum earnings, suggesting that the former is more important in determining migration choices. Regarding the nonpecuniary attributes, OLS estimates are always slightly larger than LAD estimates (especially for the study-job the match attribute), which reveals that individuals who expect the highest gains across alternatives in these attributes also report more extreme probabilities of migrating, i.e., they are more certain that they will *not* stay. The positive and at 1 percent statistically significant coefficient on the probability of having a job that is directly related to their bachelor's degree shows that expected gains in this career-related outcome is a relevant determinant of expected migration choices, even after controlling for expected earnings gains. As for nonpecuniary factors, enjoying the quality of social life and being close to family, partner and friends are positive and statistically different from zero at the 1 percent level. Overall, results show that students consider pecuniary and nonpecuniary outcomes when deciding which migration alternative to choose.

Parameter estimates are easier to interpret in terms of implied choice elasticities, willingness-to-pay estimates and changes in migration choices that result from students' belief manipulation. I present these results in the sections below. In all these sections, I show the results for the LAD estimates reported in Column (1) in Table 1.7, which refer to the utility specification that includes expected minimum earnings.

Choice elasticity of labor market outcomes

This section examines what the model estimates imply about the responsiveness of expected migration choices to changes in career-related outcomes. For example, in response to an increase in beliefs about minimum earnings in migration alternative m, how likely would an individual be to choose that migration alternative? How much more likely if, instead, for the same level of beliefs about earnings, beliefs about the probability of working full-time or part-time increased?

Table 1.8 reports migration choice elasticities using the model parameter estimates reported in the first column of Table 1.7. For each student i, I compute the percent change in the likelihood of choosing migration alternative m when beliefs about a given career-related outcome increase by 1% in each year of migration alternative m and are held constant in the other two alternatives. These choice elasticities are heterogeneous across students, as they depend on individual specific beliefs $G_{i,\tau}(x|m,t)$ about outcomes for each alternative.

Table 1.8 Elasticity: Average percent change in the likelihood of choosing a migration alternative given a 1% increase of a given outcome in every period of that alternative

	Mininimum earnings	Full-time employment	Part-time employment	Study-Job match
% Δ No-migration	0.63	0.53	0.10	0.55
% Δ Short-term migration	0.75	0.64	0.11	0.58
% Δ Long-term migration	1.01	0.88	0.13	0.74

The mean elasticity (averaged across the three migration alternatives) is 0.80 for minimum earnings. This number is higher but similar in magnitude to other recent studies that analyze migration in Spain using aggregate data and alternative identification strategies and populations (see Melguizo and Royuela (2020), Clemente et al. (2016)). As examples of results from other contexts, who also use other methodologies, the elasticity I estimate is in the range of elasticity of the choice to changes in earnings estimated by Dahl and Sorenson (2010) for Danish engineers (whose estimates are between 0.5-1%) and is somewhat lower than that estimated by Bertoli et al. (2013) for migrations from Ecuador to Spain (which is equal to 0.95%). The elasticity of choice that I estimate for the probability of working full-time is equal to 0.68. That is, individuals respond not only to higher expected earnings but also to the expected probability of having a full-time job on its own. While responsiveness is highest for minimum earnings, results suggest that both outcomes play an important role in migration choices. The elasticity of choice to increases in the probability of working part-time is much lower, equal to 0.11. Finally, the mean elasticity of the study-job match is equal to 0.62, showing that young adults expected migration choices respond to good study-job matches, after controlling for earnings. The fact that percent changes are always biggest for long-term migration is driven by the fact that, on average, individuals report a lower likelihood of choosing this alternative.

Willingness-to-pay to increase nonpecuniary attributes

Coefficients of nonpecuniary factors are easiest to interpret in terms of willingness-to-pay (WTP) estimates. For example, how much of their expected earnings are students willing to forgo to increase the probability of enjoying the quality of their social life by Δ percentage points, other things being equal? In this Section, I translate the magnitudes of the parameter estimates for enjoying the quality of social life and enjoying being close to family, partner and friends reported in Column (1) of Table 1.7 into willingness-to-pay estimates. Based on the utility specification in equation (1.5) the per-period willingness-to-pay to experience an increase equal to Δ percentage points in outcome x_j in a each period is computed as

$$WTP = \frac{\phi_j}{\phi_1}\Delta$$

where $j = \{3, 4\}$. Column 1 of Table 1.9 reports the WTP to increase the nonpecuniary factor from its expected level in the long-term migration alternative to its expected level in the no-migration alternative. Panel A reports WTP estimates for increases in the probability of enjoying the quality of social life and Panel B for increases in the probability of enjoying being close to loved ones. The first row of each panel presents the estimates in euros. Standard errors, in parenthesis, are calculated using the delta method. The second row of each panel shows the estimates as a fraction of the average -across individuals and alternatives- annual minimum earnings, where annual is calculated as the total sum of minimum earnings in each path divided by the length of the path of each individual. This number is equal to 16,322€.[16]

Table 1.9 Willingness-to-pay to increase given attribute by Δ

	Δ	
	From m_3 to m_1	1 s.d.
	(1)	(2)
	A. *Enjoy social life*	
WTP (€, year)	5,980.03***	7,080.27***
	(1665.67)	(1,972.13)
WTP (as % of avg. annual earnings)	37%	43%
	B. *Enjoy being close*	
WTP (€, year)	6,204.93***	3,527.20***
	(2320.23)	(1,318.94)
WTP (as % of avg. annual earnings)	38%	22%

Note: Standard errors in parenthesis, calculated using the delta method. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. The average annual minimum earnings: 16,322€. 1 s.d. equals 0.25 p.p. for enjoying the quality of social life and 0.32 p.p. for enjoying being close to loved ones.

[16]For reference, the average annual minimum earnings if working full-time are equal to 21,189€.

I find that students are willing to give up 37% of their average annual minimum earnings for such an increase in the chance of enjoying their quality social life. On the other hand, students are willing to give up 38%, to increase the likelihood of enjoying being close to family, partner, and friends. This number lies in the range of the WTP in order to live close to family estimated by Koşar et al. (2021) for US residents. They find an average WTP as a percent of income equal to 30% for individuals who define themselves as mobile and 56% for those define themselves as rooted . Our estimate being closer to the former than the later is consistent with young adults in our sample being less likely to be of a “rooted" type (i.e., they are less likely to own a house or have a family). Taking both nonpecuniary factors into account, young adults are willing to give up 74% of their lifetime earnings (12,185€ annually) to increase each of the nonpecuniary factors from their expected levels in the long-term migration alternative to their expected levels in the no-migration alternative. Overall, the total WTP for nonpecuniary factors is lower than the total moving costs typically estimated in migration studies, which exceed 100% of income (Kennan and Walker, 2011; Ransom, 2022).[17]

Column 2 reports the WTP estimates for increases equal to one standard deviation in each nonpecuniary outcome. This measure allows to compare the WTP of one nonpecuniary factor with the other, as it takes their dispersion into account. Results show that students have a higher WTP for the quality of social life than for being close to family, partner and friends. They are willing to pay 43% of their average annual earnings to increase the probability of enjoying the quality of social life by 1 standard deviation, while are willing to give-up 22% of their earnings for an equal increase in the probability of being close to loved ones.

The relative role of pecuniary and nonpecuniary factors

So far, we have seen that both labor market outcomes and nonpecuniary factors are statistically significant and economically important determinants of migration choices. The goal of this

[17]These studies estimate dynamic choice models and identify moving costs by assigning a distinct status to each person’s place of birth. Due to different modelling assumptions the numbers are hard to compare, but the discrepancy is likely due to two main factors. First, I am able to identify moving costs better. Second, I estimate costs for each individual’s chosen migration destination relative to staying. In the two mentioned studies, the moving cost represents the cost faced by the average individuals if they were forced to move to an arbitrary location. As the authors say, the estimated cost would be lower if individuals were allowed to choose the best available location to them.

section is to provide a meaningful metric to understand the *relative* role of the set of career-related outcomes and nonpecuniary factors on migration choices of highly-educated young adults in regions with poor labor market conditions. For this, I use the parameter estimates reported in the first column of Table 1.7 and predict the choice behavior under the two counterfactual scenarios. The first one equalizes labor market outcomes across alternatives and sets them equal to the no-migration alternative. That is, it assumes that young adults expect the same career prospects if they were to stay than if they were to follow short- or long-term migration. The second one sets the nonpecuniary factors equal across alternatives and to the no-migration alternative. In other words, it assumes that students believe their quality of social life and enjoyment from being close to loved ones will be unaffected by their migration choices. Given the described expectations and estimated preferences, we expect the first counterfactual to push students away from the long-term migration and into the no-migration. This is because nonpecuniary factors now play a larger role in sorting and because students' beliefs about labor market outcomes are much lower in the no-migration alternative than in the long-term migration alternative. Following the same logic, we expect the second counterfactual to cause the opposite. That is, to push students away from the no-migration and into the long-term migration. Because the short-term migration can be seen as a milder version of the long-term migration, we expect the inflows and outflows to have the same direction as the long-term migration but with lower magnitudes. This is precisely what I find. The question then is to investigate which counterfactual causes larger changes in the expected choice probabilities, as this gives a meaningful metric to understand the implied magnitudes of our estimates.

Results are shown in Table 1.10. The first column displays the baseline probability of choosing each migration alternative. Columns (2a) and (2b) show the percentage point increase/decrease in the probability of choosing each migration alternative under the counterfactual scenario relative to the baseline. Columns (3a) and (3b) present the mean percent changes by alternative. Note that the effects of these counterfactuals are generally heterogeneous across students.

Results show that expected nonpecuniary factors play a larger role in determining expected migration choices. In other words, the percentage point changes that result from equalizing nonpecuniary factors across alternatives are always larger than those resulting from setting labor market outcomes equal. For example, if students believed that they would enjoy the same

Table 1.10 Percent change in the probability of choosing migration alternative m when beliefs about labor market outcomes and nonpecuniary factors are equalized across alternatives, setting equal to the no-migration alternative.

	Base	Equal beliefs on career outcomes		Equal beliefs on nonpecuniary outcomes	
	Level	Δ	$\% \ \Delta$	Δ	$\% \ \Delta$
	(1)	(2a)	(3a)	(2b)	(3b)
No-migration	0.360	0.089	24.72	-0.148	-41.11
Short-term migration	0.394	-0.033	-8.37	0.038	9.64
Long-term migration	0.246	-0.057	-23.17	0.111	45.12

nonpecuniary factors if they chose the short-or long-term migration as if they were to stay, the probability of staying would decrease by 41%. Instead, if they thought their region of birth offered the same career prospects as their chosen migration destination over the life-cycle, the probability of staying would increase by 25%. Overall, results show that expected migration choices are more responsive to expected differences in nonpecuniary factors than to expected differences in labor market outcomes.

The role of planned short-term migration

This section investigates the drivers of planned short-term migration using two counterfactual exercises. Many migrations are temporary. Using administrative data, Table A.1 shows that there are more than twice as many temporary migrants as there are long-term migrants among individuals with tertiary education who move from Andalusia to the center-north of Spain. Results in section 1.4 showed that short-term migration is on average seen as 40% more likely than long-term migration by young adults in our sample. Given the its high prevalence, it is interesting to understand why young adults plan to migrate short-term.

I perform two counterfactuals, which manipulate students' expectations in the periods when individuals are back in their region of birth after having migrated temporarily, $t \in [7, 10]$. The goal of the first counterfactual is to explore the role of short-term migration as a human capital acquisition strategy. For this, it assumes that each student expects, in the short-term migration alternative in periods $t \in [7, 10]$, the same labor market outcomes that she expects in the no-migration alternative in these same periods, all else equal. The literature has often modeled temporary migration as an optimal life-cycle investment (Dustmann et al., 2011;

Dustmann and Görlach, 2016; Dustmann and Weiss, 2007; Thom, 2010), where individuals can acquire abroad, more efficiently than at home, human capital that is highly valued at home. The young engineers in our sample, for example, could plan to work in a big, internationally renowned firm in Madrid for some years, and expect that experience to help them get a better job in Andalusia some years after. Results presented in section 1.4 showed that on average this is the case. The mean of the expected *full-time* wage premium after return is equal to 13%, the mean of the expected wage premium after return is equal to 19% and the mean of the expected study-job match premium is equal to 9% (these results are presented in Table A.5).[18] This, however, does not mean that anticipating these gains is a driver of their decision-making. For example, it could be that those who expect highest premium after return are those who have most positive beliefs about the benefits of migration in general, and are those who are most likely to migrate long-term. The aim of the second counterfactual is to assess the extent to which short-term migration is motivated by individuals' willingness to avoid having too low nonpecuniary factors for too long. To this end, it assumes that each student expects, in the short-term migration alternative in periods $t \in [7, 10]$, the same nonpecuniary outcomes that she expects in the long-term migration alternative in these same periods, all else equal.

Given that students on average expect a premium after return and expect the long-tern migration to have the lowest nonpecuniary factors, both counterfactuals make the short-term migration a less appealing alternative for the average student. If these are drivers of expected short-term migration, we expect the mean expected choice of short-term migration to decrease under each of the two counterfactual scenarios relative to baseline, where individuals' expectations are not manipulated. This is what I find. The question then is to assess which of the two mechanisms drives the largest changes in the expected choice of short-term migration.

Results are presented in Table 1.11. Columns (1a) and (1b) show the percentage point change caused in the expected choices of short-term migration by counterfactual 1 and counterfactual 2, respectively. Columns (2a) and (2b) present the percent changes.

Using the same model estimates as reported in Table 1.7, I find that the choice of short-term migration would fall by 1.8 p.p. (4%) if students believed that there is a wage premium equal to 0 after return in all labor market outcomes. The fact that the change is negative shows that

[18]Expected wages refer to expected wages conditional on employment status averaged by employment status probabilities.

Table 1.11 Percent change in the probability of choosing short-term migration when beliefs are equalized for the periods when the individual is back in her region of birth *after* having migrated.

	Counterfactual 1: Equal beliefs about pecuniary outcomes ST = No migration		Counterfactual 2: Equal beliefs about nonpecuniary outcomes ST = Long-Term m.	
	Δ	$\% \Delta$	Δ	$\% \Delta$
	(1a)	(2a)	(1b)	(2b)
Short-term migration	-0.018	-4.45	-0.068	-17.45

Note: Expected probability of short-term migration at baseline is 0.394.

the anticipation of a wage premium after return influences the decision to plan a temporary migration. However, its role is rather small. Results show that changes in expected choices of short-term migration are larger under the second counterfactual. If students believed that nonpecuniary factors after having returned back to their region of birth were equal to those they would experience under long-term migration, the expected choice of short-term migration would drop by 6.8 p.p., a drop equal to 17%.

These results suggests that young individuals' choice of whether to return back to their region of birth is more sensitive to nonpecuniary conditions at the migration destination (e.g., whether they make friends, find a partner or adapt to the lifestyle of the destination) than to labor market conditions at home.

1.5.3 Predictive Validity

The employed approach emphasizes that it is the beliefs before the choice is made -not realized outcomes in later periods- which are fundamental to understand choices. Whether these perceptions are biased is then arguably an irrelevant issue from the perspective of understanding the decision.[19] The methodology instead rests on the implicit assumption that individuals' reported choices in the survey are reflective of what respondents would do in actual scenarios. There is growing evidence that individuals' reported subjective choice probabilities strong and positively correlate with their realized choices years later. See Arcidiacono et al.

[19]To the extent that biased beliefs can be corrected with information interventions, however, assessing the accuracy of beliefs is relevant from a policy-making perspective.

(2020) for occupational choices, Wiswall and Zafar (2021) for educational choices, and most relevant to this study, Koşar et al. (2021) for migration choices.

Given that I carried out the first survey in June 2020, when students were about to finish their bachelor's degree, and some students will pursue further studies before entering the labor market, the follow-up survey should be carried out in the near future. As suggestive evidence of the predictive validity of their reported choices, I look at students self-selection into migration based on individual characteristics. The finding that those who report higher migration likelihoods in my sample share the same characteristics of actual migrants in Spain would provide a first suggestive evidence that their reported expected choices are reflective of actual choices in the future. Results of Table 1.12 show that this is the case. Being male, from a high socioeconomic status, older and having higher grades are all positive and statistically significant determinants of expected migration probabilities. Interestingly, the comparison of columns 1 and 2 show that selection into migration in terms of gender is driven in part by selection into field of study. Engineering students, who are more likely to be male, are almost 60% more likely to migrate than students of Social Science and Law. Overall, results are in line with self-selection into migration of young adults in Spain (see González-Leonardo et al. (2022)).

Table 1.12 Fractional logit estimates

	Expected probability of migrating (1)	(2)
Female	0.838*	0.889
	(0.0884)	(0.0943)
High SES	1.488***	1.381***
	(0.156)	(0.148)
Age	0.936**	0.930**
	(0.0281)	(0.0273)
Above median standardized GPA	1.289**	1.312**
	(0.139)	(0.141)
Field of study		
- Engineering		1.573***
		(0.178)
- Natural science		1.165
		(0.231)
N	490	490

The outcome variable is the probability of migrating, which pulls the probability of migrating short- and long-term together. I use the with fractional logit estimator (Papke and Wooldridge, 1996) which is adequate when the outcome variable is a fraction/ probability. The table reports exponentiated coefficients. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Baseline categories are Male, Low SES (none of the parents has tertiary education) and Social Science and Law. GPA is standardized by field of study and university.

1.6 Conclusion

Young adults' migration choices at the beginning of their professional careers play a key role in determining their employability and future earnings profiles. This paper investigates the role of expected pecuniary outcomes (earnings, employment and study-job match) and nonpecuniary factors (enjoying being close to family, partner and friends and enjoying the quality of social life) on migration choices. I focus on the most mobile group of people i.e., young and highly educated individuals, and study expected migration choices out of Andalusia, one of the poorest regions of Spain. Spain is a country with large employment and income disparities across regions, similar to other European countries. I find that both expected labor market returns and nonpecuniary factors are statistically and economically significant determinants of migration choices. However, counterfactual exercises show that the relative role of the set of nonpecuniary outcomes is larger than that of labor market outcomes on young adults' out-migration choices. Additionally, given their preferences and beliefs at the time of out-migration, counterfactual exercises suggest that their choice to return back to their region of birth is also likely to depend to a larger extent on the nonpecuniary experiences at the destination, such as making new friends or forming a family, than on the labor market conditions at home. The important role played by nonpecuniary factors even for the most mobile group of people, showing that moving costs have an important social dimension, suggests a role for place-based policies as a tool to foster convergence across regions (Bartik, 2020).

2

Understanding the self-selection into internal and international migration of young individuals: Evidence using subjective expectations data

Abstract This paper studies the self-selection intentions into internal and international migration of young individuals in their school-to-work transition from the South of Spain -one of the areas with highest youth unemployment rates in the EU. I use a rich dataset that includes personal, academic and family background characteristics, as well as individuals' beliefs about labor market outcomes in their home region and migration destination. I find that individuals who plan to migrate internationally come from the highest end of the grade distribution and are from more privileged family backgrounds relative to the other two groups. Despite being positively selected, students who plan to migrate internationally have the most pessimistic views about their career prospects in their home region. With their migration plans, they expect higher labor market returns to migration than internal migrants. International migrants are more likely than internal migrants to plan a long term migration as opposed to a temporary

migration. If individuals follow their plans, my results suggest a future brain drain from the region as well as from the country.

2.1 Introduction

The choices that people make at the beginning of their careers can significantly impact their future job prospects. For instance, taking a job for which they are overeducated or working for a small company instead of a large one can limit their ability to gain valuable experience and skills, which can negatively affect their long-term career prospects (e.g., Arellano-Bover, 2020, 2022; Baert et al., 2013). Young adults who restrict their job search to areas close to their birthplace may be more likely to make such choices, specially when these areas have limited economic opportunities. Therefore, policymakers are interested in creating platforms such as job portals or counseling services to assist young adults in finding suitable job matches across broad areas at the beginning of their careers, whether within or across countries. One such example is the EURES platform, which facilitates the free movement of young workers across European countries. At the same time, governments in given countries are keen on knowing the characteristics of young adults who sort into each type of migration, as selective out-migration to other regions and other countries have different compositional effects on the workforce that spreads across the country.

In this paper, I study future migration plans of individuals in school-to-work transition, distinguishing among those who plan to move internally and internationally. First, I analyze self-selection into staying, migrating to another region and migrating to another country in terms of a rich set of individual characteristics. Then, I look at group differences in expectations regarding own labor market prospects and personal life in their region of birth. Finally, I calculate each groups' expected returns to migration, and study how much of the differences across groups are explained by differences in group composition in terms of individual characteristics. To do so, I collected data from 609 young adults who were about to finish the bachelor's degree in the most populous region of Spain, which has one of the highest youth unemployment rates in Europe (Andalusia). Given large regional inequalities within Spain but also the fact that labor market prospects are better in many other advanced economies that surround the country, young adults born in Andalusia have economic incentives to sort both into internal as well as international migration.

The findings indicate that individuals who anticipate migrating, either internally or internationally, have higher expected levels of education, are more likely to be pursuing STEM degrees, and are younger compared to those who intend to remain in their birth region. However, there are no significant differences between the two groups of migrants on these aspects. A distinct trend emerges regarding selection based on their academic performance, as the grade distribution of international migrants first-order stochastically dominates the grade distribution of both stayers and internal migrants. Additionally, results show that both internal and international migrants hail from more privileged family backgrounds than non-migrants, with international migrants being particularly advantaged. For simplicity, I often refer to individuals who *plan* to stay, to move internally and to move internationally as “stayers", “internal migrants" and “international migrants", but it should be noted that the terms refer to their plans to belong to these groups as opposed to actual realizations.

The second set of results analyzes how the three groups differ in terms of beliefs about life if they were to stay in their region of birth. Differences between internal migrants and stayers arise from factors beyond the labor market, as stayers exhibit higher probabilities of enjoying the quality of social life and enjoying being close to loved ones than internal migrants if they were to stay at birthplace. On the contrary, both groups anticipate comparable labor market outcomes if they were to stay in their region of birth. In contrast, international migrants and stayers diverge primarily in labor market prospects, as international migrants express notably more pessimism about full-time job opportunities and have higher earnings uncertainty in their region of birth, despite being positively selected based on education and family background characteristics. As a result, the gap in pessimism is exacerbated when we control for observed characteristics. International migrants are also significantly more likely to believe that they will struggle financially and that they will not have enough money to do the things that they enjoy if they were to stay in their birthplace. Compared to realized labor market outcomes of comparable individuals in Andalusia, all groups have pretty accurate beliefs about mean wages if working full-time in their region of birth. However, all groups report lower chances of working full-time and higher chances of working part-time than full-time and part-time employment rates of young individuals with tertiary education in Andalusia, particularly so international migrants.

Third and finally, I study differences in expected returns to migration across the three migrant groups, and decompose the gaps into parts that can be explained by the different individual characteristics. Results show that international migrants expect significantly higher returns to migration both in labor market outcomes as well as in personal life related outcomes relative to stayers. The higher expected returns in both dimensions are largely explained by their higher likelihoods of having networks abroad and to a lesser extent, bu their socioeconomic status.

This paper contributes to the migration literature by analyzing self-selection intentions into internal and international migration of a population with incentives for both. Due to a lack of comprehensive datasets that allow to integrate both type of migrants in a single framework, the literature has typically studied internal and international migration separately (some exceptions include Gröger (2021)). However, in order to understand the allocation of talent, it is important to understand who and why chooses to move internally and internationally. This study overcomes this issue by including in a common survey design potential future international, internal and non-migrants.

The paper contributes to and builds on two strands of the literature. On the one hand, to the large literature studying self-selection into international migration (e.g, Abramitzky et al., 2012; Belot and Hatton, 2012; Borjas et al., 2019; Chiquiar and Hanson, 2005; Fernández-Huertas Moraga, 2011; Grogger and Hanson, 2011; Kaestner and Malamud, 2014; McKenzie et al., 2010; Patt et al., 2021; Rosso, 2019) and self-selection into international migration within the highly educated individuals (e.g., Albarrán et al., 2017; Barone et al., 2019; Gibson and McKenzie, 2011b). On the other, to the literature on self-selection into internal migration (e.g, Barone et al., 2019; Bartolucci et al., 2018b; De la Roca, 2017). I complement this literature by using direct measures of several educational outcomes that the typical dataset does not contain, such as grades, and study self-selection intentions based on how good or bad, relative to one another, different groups expect to perform at birthplace as opposed to how good or bad they were actually performing before choosing to migrate.

The paper is organized as follows. Section 3.3 describes the data. Section 2.3 the belief elicitation in the survey and construction of measures. Section 2.4 describes self-selection intentions into staying, migration internally and migrating internationally by individual characteristics. Section 2.5 reports the differences across migrant groups in expected about labor market and personal life outcomes if they were to stay in the birthplace and Section 2.6 compares these beliefs to actual realizations of a comparable population. Section 2.7 analyzes differences in expected returns across the migrant groups.

2.2 Description of data

This section describes the institutional setting, survey administration and sample characteristics.

2.2.1 Institutional Setting

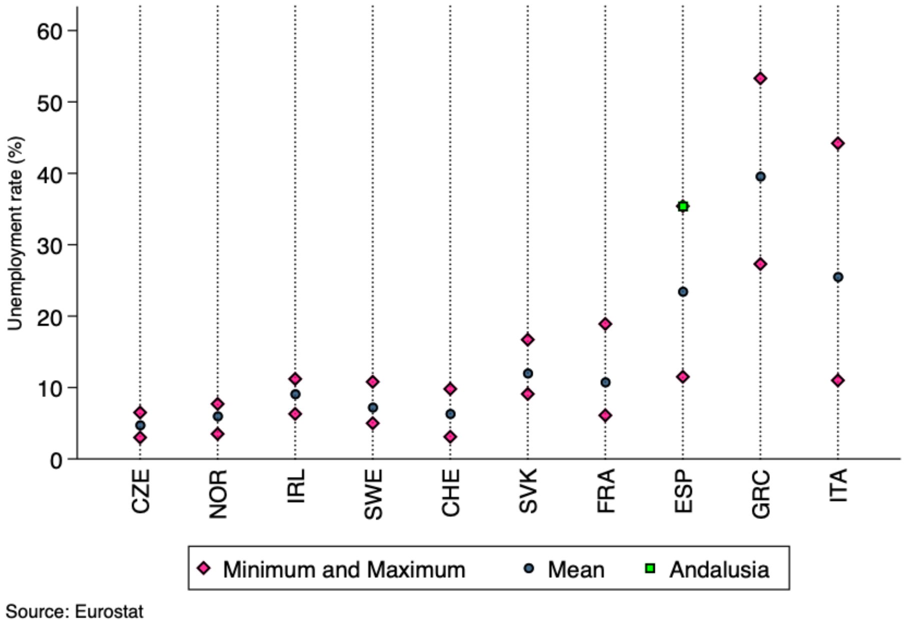
I gathered a sample of last year bachelor's degree students born in the region of Andalusia in Spain. Andalusia is the most populous and second largest region in the country (with over 8 million inhabitants and over 87,000 km^2). Geographically, it is the southernmost region of the country. The south of Spain has been historically poorer, with persistently lower income and employment rates. To this day, per person net income in the south is about 60% of that in the center-north, and unemployment rates are around double. For being the most populous region of the country, the south is often represented by Andalusia.

As shown in Figure 2.1, Andalusia is the region with the highest unemployment rate of young individuals with tertiary education in the country, more than 3 times higher than the region with the lowest unemployment rate for this group in Spain. As shown in the figure, however, young and highly educated adults in Andalusia can also largely benefit from searching for jobs in other countries altogether, as in many surrounding countries unemployment rates for this group are considerable lower than those in Spain.

2.2.2 Survey Administration

The surveys were administered in the two biggest universities of the region of Andalusia: the University of Seville (US) and the University of Granada (UGR). The study was limited to the Department of Economics and Business, the Department of Engineering and the Department of Natural Sciences. The data collection was carried out in two consecutive surveys, 3 days

Figure 2.1 Youth unemployment rate by regions in European countries



Source: Minimum, maximum and mean youth unemployment rates (15-29 years old) of young adults with tertiary education in European countries, year 2016.

apart. Both surveys were constructed using Qualtrics Survey Software and students received the link to both surveys via email.

The first survey was administered directly by the universities to a total of 5.296 last year bachelor's degree students. The link was closed less than 12 hours after it was activated, when we reached the sample size that our budget limit allowed to compensate. The take-up rate during this time was 18.07% (957 completed responses). This survey collected data on individual and family background characteristics and took, on average, 7 minutes to complete. Participation in the survey was encouraged through a game of chance, where students could win real money, ranging from 0€ to 7.5€ with varying probability. The purpose of this survey was twofold: First, to collect students' contact information (phone number and email address) so that researchers could contact students directly in the next round, without the intermediation of the university. Students were explained that the email address would be used to send them the links to future surveys, and that the phone number would be used to make survey payments -via a widely used phone payment platform in Spain (Bizum)-, as well as to notify them about the sending of surveys. Second, the survey served to collect data on students' place of birth and only students who were born in Andalusia were invited to participate in the second survey. Students were not told the selection criterion to receive the second link and were warned that their participation in the second survey was not guaranteed. 688 students -those born in Andalusia- were sent the link to the second survey.

Students were given 3 weeks to submit the second survey and they were told to complete the survey in one sitting. The second survey took, on average, 30 minutes to complete, and each student was compensated with 6€, which were paid by phone via Bizum. This survey collected individuals' expectations about different outcomes in their region of birth and at chosen migration destinations. Out of the 688 students that were sent the second survey, 609 completed it (89.53% follow-up rate).

2.2.3 Sample Characteristics

Table 2.1 shows the sample's descriptive statistics. It also shows how students in the sample relate, in terms of educational characteristics, to the population of students studying the same degrees in the same departments and, in terms of family background characteristics, to the population of Andalusian university students. 81% of students in the sample are in their last year of bachelor's degree and 37% are studying a STEM degree. The fraction of females, the mean age at degree completion, and the average GPA in my sample are very similar to the ones in the population of students enrolled in the last year of equivalent degrees, i.e., 46% vs. 48% of females, 23 vs. 24 years old, and 6.82 vs. 6.9 GPA in my sample relative to the administrative data respectively. The administrative data come from the Spanish Ministry of Universities and are publicly available.[1] The survey also collected data on students' family background. In particular, about parents' highest level of education, employment status, perceived family wealth relative to the average family in Andalusia, and about the place where elder siblings where working, for those who had elder siblings and these were working. 49% of students in the sample have high socioeconomic status (defined as having at least one parent with a university degree), 50% have both of their parents working, 16% have one of their parents unemployed, and 30% think that their family wealth is higher than the wealth of the Andalusian average family. To assess how my sample's students' household characteristics relate to families who

[1] See link for individual characteristics this other link for grades by university and degree.

Table 2.1 Sample characteristics and comparison to other data sources

	My sample		University and LFS data
	Mean (s.d.)	N	Mean
Last Year (%)	81	609	
STEM degree (%)	37	609	
Female (%)	46	609	48
Age at survey completion	23.09 (2.23)	609	
Expected age at end of bachelor's degree	23.30 (2.27)	609	24
GPA, 0-10 scale	6.82 (0.84)	534	6.9**
High SES (%)	49	601	46
Parents' employment status (%)			
- Both parents working	50	609	38
- One parent working	38	609	46
- Both parents unemployed	2	609	1
- One parent unemployed	16	609	14
Perceived family wealth w.r.t. the average family in Andalusia (%)			
- Higher or much higher	30	600	
- Similar	56	600	
- Lower or much lower	14	600	
Place of work of elder siblings (%)			
- Andalusia	67	234	
- Other region within Spain	18	234	
- Other country	15	234	

Note: Last year refers to the fraction of students enrolled in their last year of the bachelor's degree. Students are categorized as high socioeconomic status (SES) if at least one of their parents attended university, and low-ses otherwise.

have children attending university in Andalusia, I use the Spanish Labor Force Survey (LFS), year 2019. I define a comparable family as one that (i) is living in Andalusia and (ii) has at least one child born in Andalusia and attending university in 2019. Overall, the results show that the family characteristics of students in the sample are similar to those of Andalusian families that have children attending university that year. For example, 46% of these students in the LFS and 49% of students in my survey have a high socioeconomic status.

2.3 Belief Elicitation and Construction of Measures

This section describes survey questions that elicit individuals' expectations and construction of measures using the question responses.

A.1. Elicitation of future migration plans: The survey elicited students' migration plans in two consecutive questions. First, students were asked about their probability of migrating within the next 10 years after finishing the bachelor's degree. Second, they were asked about the place where they would migrate [another Spanish region/another country] *if* they were to migrate. This question was asked to *all* students -including those who attached 100 probability to migrating.

B.1. Construction of Migrant Types: Using the responses to the above questions, each type of migrant is defined as follows.

- **Non-migrant**: Individuals who attach less than 70% probability to migrating.
- **Internal migrants**: Individuals who attach 70% probability or more to migrating and choose another Spanish region as migration destination.
- **International migrants**: Individuals who attach 70% probability or more to migrating and choose another country as migration destination.

Using these definitions, Table 2.2 reports the number and the fraction of individuals in each of the migration categories.

Table 2.2 Migrant types and chosen destinations by migrant types

	Type of migrant		
	Non-migrant	Internal Migrant	International migrant
Number	299	138	120
Fraction	54%	25%	22%

Next, I describe questions about individuals' expectations regarding personal and labor market related outcomes. All outcomes were asked for 3 years after finishing the bachelor's degree, and students were told to assume that by this time they had finished their highest expected level of studies. For all questions, students were asked to consider the job offers that they thought they would receive and that they would accept.

A.2. Elicitation of Distribution of expected earnings: The survey first asked each student to give their expected range of earnings at home and abroad as follows:

> What do you think is the minimum and maximum monthly gross salary that you would be earning IF you were working full-time 3 years after finishing your bachelor's degree in (a) Andalusia and (b) [*student's chosen destination*]?

Then, based on the responses to the above question, for each of the two migration scenarios, each student was shown 5 adjacent earnings intervals characterized with 4 thresholds. The thresholds that individuals were shown were determined as follows: First, using a question branching algorithm in the survey, each individual was assigned 1 out of 10 possible branches. The branch that each individual was assigned to depended on the midpoint of their reported range. In particular, this midpoint could fall into one of the following 10 intervals: (1) $\leq$ 1400€; (2) (1400,1600]; ... ; (9) (2800,3000]; (10) $>$ 3000€. Each interval corresponded to a branch which had 4 predetermined thresholds. For example, assume that a student reported to expect to earn a minimum of 900€ and a maximum of 1500€ if working full-time in her region of birth. The midpoint of this range belongs to interval (1). Then, this student, together with all students whose midpoint was lower than or equal to 1400€, were asked about the percent chance that their wage would be in each of the following 5 intervals: (1) $\leq$ 1000€; (2) (1000€,1200€]; (3) (1200€ ,1400€]; (4) (1400€ ,1600€]; (5) $>$ 1600€. Students reported the percent chance of each interval using a sliding bar, which was set at zero. Using sliding bars has been suggested as a method to avoid bunching e.g. Alesina et al. (2018)). Instead of eliciting the information in the form of a cumulative distribution (cdf), as in Dominitz and Manski (1997), I used a probability density (pdf). Experimental evidence suggests that individuals find assessing the probabilities that the outcome lies in each interval less demanding than assessing the probabilities that the outcome does not exceed the sequence of thresholds Bover (2015).

A.3 Elicitation of Employment status probability: The survey asked the following question:

> What do you think is the percent chance of
> [Working full-time/working part-time/being unemployed/being out of the labor market]
> if you were living in (a) Andalusia and (b) [*student's chosen destination*]?

B.2. Construction of Expected weighted mean earnings: This is defined as the weighted sum of expected earnings conditional on employment status where the weights are the employment status probabilities. Note that because the survey only asked about earnings conditional on working full-time, as in Wiswall and Zafar (2015), I assume that earnings conditional on working part-time are half of those if working full-time. $E(w_i) = Pr(FT) * E(w_{i,FT}) + Pr(PT) * (E(w_{i,FT})/2)$

B.2. Construction of Expected earnings gap: Earnings gap is defined as the difference between expected mean earnings conditional on working full-time and expected unconditional mean earnings. i.e., $E(w_{i,FT}) - E(w_i)$

A.4. Elicitation of Job-Study match probability: What do you think is the percent chance that you would be working in a job directly related to your bachelor's degree studies if you were living in (a) Andalusia and (b) [*student's chosen destination*]?

A.5. Elicitation Elicitation of Network to help finding a job: What do you think is the percent chance that you will have a network that will help you find a job in (a) Andalusia and (b) [*student's chosen destination*]?

A.6. Elicitation of Personal situation: What do you think is the percent chance that (1) you will enjoy the quality of social life, (2) you will enjoy being close to family, partner and friends in (a) Andalusia and (b) [*student's chosen destination*]?

A.7. Elicitation of Economic situation: What do you think is the percent chance that (1) you will struggle economically, (2) you will have enough money to do what you enjoy in (a) Andalusia and (b) [*student's chosen destination*]?

2.4 Selection in terms of education, family background and other individual characteristics

The next tables describe how internal migrants and international migrants differ from non-migrants and from one another in several education characteristics (Table 2.3) and family backgrounds and networks (Table 2.4). For binary variables, the mean value represents the fraction of students with that given characteristic among non-migrants, internal migrants and international migrants. Columns (3) and (5) report respectively the difference in means and the significance level of a test of equality in means between internal and international migrants relative to non-migrants. Column (6) reports the analogous comparison for international migrants relative to internal migrants.

Table 2.3 shows that the fraction of students who plan to pursue further studies after finishing the bachelor's degree is significantly higher among the two migrant groups relative to the non-migrant group. These fractions however do not differ between internal and international migrants. With respect to the field of study, the fraction of individuals studying in STEM fields is significantly higher among internal and international migrants relative to non-migrants. Again, the fraction between the two migrants does not differ one from another. The table also shows that the fraction of women is lowest among internal migrants and that the average age is ighest among non-migrants.

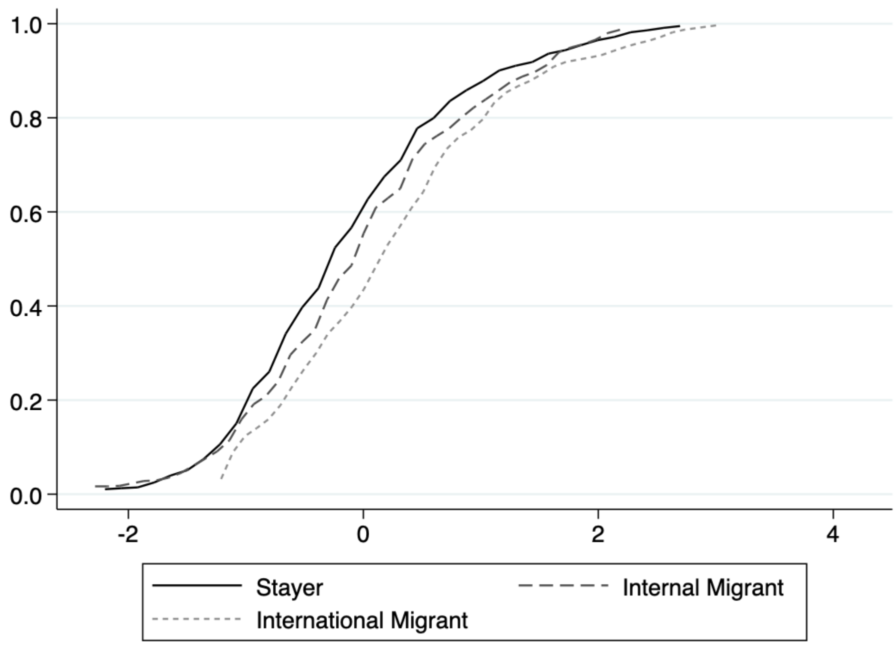
Table 2.3 Selected characteristics by expected type of migrant

	Non-migrant	Internal Migrant		International Migrant		
	Mean (1)	Mean (2)	Diff. from non-migrants (3)	Mean (4)	Diff. from non-migrants (5)	Diff. from internal migrants (6)
A. EDUCATION & PERSONAL CHARACTERISTICS						
> Bachelor's	0.82	0.90	0.08**	0.89	0.07*	-0.01
	(0.38)	(0.30)		(0.31)		
STEM	0.31	0.44	0.13**	0.43	0.12**	-0.01
	(0.47)	(0.50)		(0.50)		
Female	0.49	0.39	-0.10**	0.44	-0.05	0.05
	(0.50)	(0.49)		(0.50)		
Age	23.32	22.78	-0.54**	22.57	-0.74***	-0.20
	(2.38)	(1.87)		(1.89)		

-

The fact that internal and international migrants do not differ in terms of highest expected level of studies and likelihood of studying STEM fields contrasts with a neat pattern in grades distribution across the three groups. Figure 2.2 shows that for any cumulative probability value, international migrants always have a higher GPA than the other two groups. In other words, the the cumulative distribution functions (c.d.f.) of the international migrants first order stochastically dominates that of internal migrants and non-migrants, the difference being statistically significant at the 1% level between international migrants and non-migrants.

Figure 2.2 CDF of standardized average GPA by expected type of migrant



Note: Average GPA in previous academic year standardised by university and field of study. p-values from a Kolmogorov-Smirnov test of equality of distributions between internal migrants and non-migrant $p = 0.42$ and between international migrants and non-migrant $p = 0.009$

Table 2.4 shows how the fraction of students differs among the migrant groups by family backgrounds and networks. The fraction of students with a high SES increases from non-migrants (41%) to internal migrants (53%) to international migrants (68%). These differences are not only statistically but also economically significant. The fraction of individuals with a high SES is 29% higher among internal migrants than among non-migrants and 66% higher among international migrants than among non-migrants. In this case, differences between internal and international migrants are statistically significant too. The fraction of students reporting having higher relative wealth is statistically significantly higher for internal and international migrants relative to non-migrants.[2] Finally, while the fraction of father's employment status is similar across the migrant groups, differences stem by mother's employment status. In particular, the fraction of students whose mother working is significantly higher for internal and international migrants relative to non-migrants. Taken together, these results show that migrants, and in particular international migrants, are from more privileged family backgrounds.

To finish with, Table 2.4 analyzes differences in migrants' exposure to migration as measured by elder siblings' working locations and networks. The fraction of elder siblings working in Andalusia is statistically significantly lower for international migrants relative to non-migrants (27%), and the fraction of elder siblings working in another country is statistically significantly higher for international migrants relative to non-migrants (200%). Finally, the fraction of elder siblings working in another Spanish region is not statistically significant across the groups. International migrants perceive significantly lower chances to have networks that can help them find a job at birthplace relative to non-migrants, and internal migrants to have significantly higher networks than non-migrants and international migrants at their migration destinations.

[2]The survey asked students about what they perceived the wealth of their family was [higher/probably higher/similar/probably lower/lower] relative to an Andalusian average family. I construct an index where the "similar' category takes value equal to 0, probably higher/lower categories take the values 1 and -1 respectively and higher/lower categories take the values 2 and -2 respectively.

Table 2.4 Selected characteristics by expected type of migrant

	Non-migrant	Internal Migrant		International Migrant		
	Mean (1)	Mean (2)	Diff. from non-migrants (3)	Mean (4)	Diff. from non-migrants (5)	Diff. from internal migrants (6)
B.FAMILY BACKGROUND AND NETWORKS						
High SES (%)	0.41	0.53	0.12**	0.68	0.27***	0.15**
	(0.49)	(0.50)		(0.47)		
Relative perceived wealth (%)	0.09	0.22	0.13	0.36	0.27***	0.14
	(0.74)	(0.88)		(0.84)		
Mother's labor market status (%)						
- Out of labor mkt	0.29	0.24	-0.05	0.23	-0.06	-0.01
	(0.45)	(0.43)		(0.42)		
- Unemployed	0.14	0.10	-0.04	0.09	-0.05	-0.01
	(0.35)	(0.31)		(0.29)		
- Working	0.57	0.66	0.09*	0.68	0.11**	0.02
	(0.50)	(0.48)		(0.47)		
Father's labor market status (%)						
- Out of labor mkt	0.14	0.11	-0.03	0.17	0.03	0.06
	(0.35)	(0.32)		(0.38)		
- Unemployed	0.08	0.07	-0.01	0.05	-0.03	-0.02
	(0.27)	(0.25)		(0.22)		
- Working	0.78	0.82	0.04	0.78	-0.01	-0.04
	(0.41)	(0.39)		(0.42)		
Place of work of elder siblings (%)						
- Andalusia	0.72	0.68	-0.04	0.52	-0.19**	-0.16
	(0.45)	(0.47)		(0.51)		
- Other region	0.17	0.19	0.02	0.14	-0.03	-0.05
	(0.38)	(0.39)		(0.35)		
- Other country	0.11	0.13	0.02	0.33	0.22***	0.20**
	(0.32)	(0.34)		(0.48)		
Network at birthplace	0.60	0.56	-0.04	0.54	-0.06**	-0.02
	(0.24)	(0.25)		(0.25)		
Network at destination	0.43	0.53	0.10***	0.47	0.03	-0.06*
	(0.27)	(0.26)		(0.28)		

2.5 Beliefs about personal life and labor market prospects at birthplace

Next, I study whether the three groups differ in terms of expectations about personal life and labor market prospects if they were to stay in the birthplace. Table 2.5 shows the results that relate to earnings prospects and table 2.6 those that relate to other labor market and personal life outcomes at birthplace.

The first row in each table shows that internal migrants and stayers do not expect significantly different labor market outcomes if they were to stay in the birthplace: Internal migrants expect a higher earnings gap (the difference between the earnings that they would expect to earn if they could work full-time and the earnings that they expect as a result of the employment probabilities at birthplace), lower earnings if working-full time, higher earnings uncertainty, lower chances of working in a job directly related to their field of study, less money to do the activities that they enjoy and higher chances to struggle financially relative to stayers, if they were to stay in their region of birth. However, none of these differences are neither statistically nor economically significant. Instead, the main differences regarding expectations at birthplace between internal migrants and stayers stem from personal life outcomes, such as the probability of enjoying the social life and the probability of enjoying being close to family, partner and fiends in their region of birth. In particular, internal migrants expect 21% and 30% lower probabilities of enjoying each of these outcomes if they were to stay in the birthplace relative to non-migrants.

The picture that emerges is different if we compare international migrants to non-migrants. In this case, international migrants expect a significantly higher earnings gap than non-migrants, which stems from their lower expected chances of working full-time and higher chances of working part-time -as opposed to lower expected wages if working full-time. The gap is equal to 73€ per month, which amounts to 12% of expected earnings at birthplace. International migrants also expect significantly higher uncertainty, as measured by the variance of the individually fitted earnings distributions, equal to 29% than non-migrants. International migrants also expect significantly lower chances to have money to do what they enjoy and significantly higher chance to struggle financially relative to non-migrants (36% and 37% respectively). Instead, international migrants and non-migrants do not differ in their views about how much they would enjoy their social life at birthplace. Finally, while they expect lower chances than non-migrants of enjoying being close to loved ones if they were to stay, the difference is lower than that expected by internal migrants.

Results thus far show that international migrants expect worse labor market prospects if they were to stay in the birthplace than the other two migrants groups. This pessimistic view emerges despite international migrants being positively selected in terms education related outcomes. As shown in table 2.3 the fraction of individuals planning to pursue further studies after finishing the bachelors' degree and studying a STEM degree is highest among international migrants, and as shown in figure 2.2, they are clearly positively selected in terms of grades. Indeed, as shown by the even columns of tables 2.5 and 2.6, controlling for individual characteristics widens the differences in expected labor market outcomes between international migrants and non-migrants, given that as expected, higher levels of education, studying a STEM degree and having higher grades are all positively related to expected labor market prospects at birthplace. In short, controlling for individual characteristics, international migrants present a very pessimistic view about labor market prospects at home. Figure B.1 in the Appendix shows the distributions of these variables for the trhee migrant groups.

The next section discusses how the beliefs about labor market outcomes that the three groups hold relate to actual labor market outcomes in administrative data.

Table 2.5 OLS estimates

	(1) Earnings Gap	(2) Earnings Gap	(3) FT Earnings	(4) FT Earnings	(5) Earnings Uncertainty	(6) Earnings Uncertainty
Internal migrant	37.29	40.08	0.0104	-0.0313	0.103	0.0815
	(30.33)	(31.04)	(0.0344)	(0.0332)	(0.120)	(0.120)
International migrant	72.97**	82.42**	0.0320	-0.0240	0.266**	0.270**
	(33.47)	(34.80)	(0.0333)	(0.0334)	(0.119)	(0.126)
Higher than bachelor's		52.38*		0.0771**		0.138
		(31.50)		(0.0309)		(0.156)
STEM		-26.82		0.0660**		-0.0206
		(26.84)		(0.0281)		(0.0993)
Average standardized GPA		-33.26**		0.00711		-0.0373
		(13.90)		(0.0150)		(0.0511)
High SES		-22.13		0.129***		-0.0141
		(26.16)		(0.0277)		(0.104)
Female		16.16		-0.123***		-0.257**
		(25.21)		(0.0262)		(0.0995)
Network at birthplace		-124.9**		0.0223		0.152
		(59.80)		(0.0582)		(0.235)
Constant	463.3***	501.7***	7.187***	7.098***	10.62***	10.55***
	(16.75)	(48.98)	(0.0189)	(0.0465)	(0.0673)	(0.207)
N	549	549	549	549	549	549
adj. R^2	0.006	0.022	-0.002	0.094	0.005	0.009

- * p<0.1, ** p<0.05, *** p<0.01. + p<0.1, ++ p<0.05, +++ p<0.01 in row of international migrant t-test of difference between internal and international migrant categories.

Table 2.6 Fractional Logit estimates

	(1) Match	(2) Match	(3) Leisure money	(4) Leisure money	(5) Struggle Fin.	(6) Struggle Fin.	(7) Social Life	(8) Social Life	(9) Enjoy Close	(10) Enjoy Close
Internal migrant	-0.125	-0.224**	-0.0148	-0.0321	-0.0556	0.0442	-0.242*	-0.221*	-0.367**	-0.240
	(0.106)	(0.104)	(0.0990)	(0.0963)	(0.110)	(0.112)	(0.129)	(0.127)	(0.158)	(0.157)
International migrant	-0.194	-0.331***	-0.443***	-0.484***	0.313***	0.451***	-0.131	-0.0906	-0.281*	-0.0988
	(0.119)	(0.124)	(0.106)	(0.108)	(0.113)	(0.118)	(0.137)	(0.142)	(0.159)	(0.163)
Higher than bachelor's		0.126		0.299***		-0.242*		0.208		-0.0945
		(0.132)		(0.114)		(0.124)		(0.152)		(0.201)
STEM		0.351***		-0.0752		-0.105		-0.120		-0.254*
		(0.0946)		(0.0847)		(0.0928)		(0.113)		(0.132)
Average standardized GPA		0.0866*		0.0555		-0.0488		-0.0778		-0.0642
		(0.0460)		(0.0412)		(0.0489)		(0.0498)		(0.0582)
High SES		0.320***		0.216***		-0.265***		0.246**		-0.0154
		(0.0923)		(0.0834)		(0.0928)		(0.111)		(0.133)
Female		-0.217**		-0.118		0.351***		-0.0136		0.253**
		(0.0868)		(0.0792)		(0.0862)		(0.110)		(0.126)
Network at birthplace		0.792***		1.130***		-0.0764		1.390***		1.971***
		(0.195)		(0.173)		(0.193)		(0.242)		(0.280)
Constant	0.277***	-0.419**	0.509***	-0.402***	-0.826***	-0.631***	1.465***	0.442**	1.921***	0.874***
	(0.0595)	(0.179)	(0.0548)	(0.150)	(0.0560)	(0.160)	(0.0765)	(0.210)	(0.0943)	(0.274)
N	549	549	549	549	549	549	549	549	549	549
adj. R^2										
pseudo R^2	0.001	0.023	0.006	0.026	0.003	0.015	0.002	0.026	0.004	0.045

- Omitted categories are: Stayer, bachelor's degree, no STEM degree, low SES and male. Standard errors in parenthesis. * p<0.1, ** p<0.05, *** p<0.01. + p<0.1, ++ p<0.05, +++ p<0.01 in row of international migrant t-test of difference between internal and international migrant categories.

2.6 Comparison of actual and expected labor market outcomes at birthplace

Table 2.7 Comparson of beliefs and realized outcomes in administrative data of comparable population

	Realized outcomes	**Beliefs by migrant groups**		
	(Data from LFS/MCVL)	Non-migrants	Internal	International
Monthly gross wage				
if working F.T. €	1463.69	1,392.96	1,414.63	1,424.51
	(700.11)	(507.95)	(540.88)	(483.93)
Individual uncertainty				
S.D. €		231.13	243.28	259.00*
		(117.63)	(118.57)	(145.55)
Employment status probabilities				
- Full-time	0.66	0.51	0.48	0.46
	[0.03]	(0.26)	(0.25)	(0.27)
- Part-time	0.12	0.28	0.29	0.29
	[0.02]	(0.17)	(0.16)	(0.18)
- Unemployed	0.17	0.17	0.17	0.20
	[0.02]	(0.16)	(0.15)	(0.18)
- Out of labor mkt	0.05	0.05	0.07	0.05
	[0.01]	(0.08)	(0.10)	(0.11)
Wage weighted by				
employment status €,		930.19	915.12	895.89
		(519.25)	(549.83)	(529.02)
Wage gap				
(full-time - weighted)		462.77	499.51	528.62*
		(286.55)	(293.82)	(318.28)
Good study/job match	0.47	0.57	0.54	0.52
	(0.49)	(0.25)	(0.25)	(0.28)

Note: S.d. in parenthesis. p-values of t-test between internal and international vs non-migrant. * p<0.1, ** p<0.05, *** p<0.01. p-values of t-test between international vs internal. + p<0.1, ++ p<0.05, +++ p<0.01. Realized outcomes data are for the year 2017 and the source is the Spanish Continuous Work History Sample (MCVL). Restricted to individuals with tertiary education, born in Andalusia and working in Andalusia who were between 26-29 years old in 2017. Salaries are proxied by social security cotization bases, whose minimum and maximum were 825.60€ and 3.751,20€ in year 2017. To be consistent with the contruction of the variable in my survey, I set unemployment benefits equal to 0. S.e. in brackets. Realized outcomes are from the LFS, 2019T1. Sample is restricted to individuals between 25-30 years old with completed university degree (currently not studying), born and living in Andalusia. Realized study-job match is from the MCVL. It is a dummy variable that takes value equal to one if the individual's contract belongs to one of the following three categories: (1) Ingenieros, licenciados y alta dirección, (2) Ingenieros técnicos peritos y ayudantes,(3) jefes administrativos y de taller.

2.7 Expected returns from migration by expected type of migrant

This section analyzes differences in expected returns to migration for the different labor market and social life outcomes. As before, I compare expected returns of internal migrants and international migrants relative to non-migrants. The table shows that international migrants expect highest returns to migration in all labor market related outcomes: They expect 20% higher gains in wages by migrating, 11% higher difference in chances of working full-time and and in chances of working in a job directly related to their field of study than non-migrants migrants. They also expect higher gains (lower penalties) in enjoying the quality of social life and enjoying being close loved ones than internal migrants. The "Explained" column decomposes the parts of this gaps that can be explained by each of the covariates, conditional on all of them simultaneously (following Gelbach (2016)). The table shows that the variation in the expected probabilities of having a network that will help them find a job and to a lesser extent, the socioeconomic status explains the largest part in the gaps between the two groups.

Internal migrants and stayers do not expect significantly different returns to migration in labor market outcomes. Instead, results suggest that their different migration choices are determined by differences in how migration will affect their personal lives, as measured by enjoying quality of social life and enjoying being close to loved ones.

Table 2.8 Expected differences in given outcomes between chosen destination and Andalusia by migrant group

	Differences in expected outcomes		
	Non-migrants	Internal	International
Δ Log weigthed earnings	0.29	0.26	0.67*** +++
	(0.39)	(0.30)	(0.50)
- Δ Log Earnings if FT	0.17	0.14	0.37*** +++
	(0.20)	(0.15)	(0.26)
-Δ Full-time	0.08	0.08	0.19*** +++
	(0.18)	(0.17)	(0.24)
-Δ Part-time	-0.01	-0.01	-0.04*
	(0.14)	(0.12)	(0.18)
-Δ Unemployed	-0.06	-0.05	-0.12*** +++
	(0.14)	(0.10)	(0.16)
-Δ Out	-0.01	-0.02	-0.02*
	(0.07)	(0.08)	(0.11)
Δ SD	-0.14	-0.13	0.10*** +++
	(0.26)	(0.26)	(0.30)
Δ Job/Study Match	0.07	0.10	0.17*** +++
	(0.19)	(0.19)	(0.25)

Table 2.9 Decomposing the internal-stayer and international-stayer in expected returns in several outcomes into individual characteristics

	Expected Returns														
	Weighted Earnings			**Full-time Employment**			**Study-Job Match**			**Enjoy Social Life**			**Enjoy being Close**		
	Specification			Specification			Specification			Specification			Specification		
	Base	Full	Explained	Base	Full	Explained	Base	Full	Explained	Base	Full	Explained	Base	Full	Explained
Internal-Stayer Gap	-0.023	-0.017	-0.006	0.007	0.008	-0.000	0.034*	0.029	0.005	0.106***	0.087***	0.019***	0.141***	0.092***	0.049***
	(0.017)	(0.017)	(0.005)	(0.018)	(0.018)	(0.005)	(0.020)	(0.020)	(0.006)	(0.024)	(0.024)	(0.007)	(0.032)	(0.032)	(0.013)
Education	No	Yes	-0.002	No	Yes	-0.004	No	Yes	-0.001	No	Yes	0.004	No	Yes	0.001
			(0.003)			(0.003)			(0.003)			(0.004)			(0.004)
Avg GPA	No	Yes	0.001	No	Yes	-0.000	No	Yes	-0.001	No	Yes	0.003	No	Yes	0.003
			(0.001)			(0.001)			(0.001)			(0.003)			(0.003)
SES	No	Yes	-0.003	No	Yes	-0.003	No	Yes	-0.004	No	Yes	0.001	No	Yes	0.007
			(0.003)			(0.002)			(0.003)			(0.003)			(0.004)
Gender	No	Yes	-0.002	No	Yes	-0.001	No	Yes	-0.003	No	Yes	-0.001	No	Yes	0.005
			(0.002)			(0.002)			(0.002)			(0.002)			(0.004)
Network Difference	No	Yes	0.001	No	Yes	0.008**	No	Yes	0.013***	No	Yes	0.013**	No	Yes	0.032***
			(0.003)			(0.004)			(0.005)			(0.005)			(0.010)
International-Stayer Gap	0.199***	0.207***	-0.008	0.115***	0.122***	-0.006	0.112***	0.115***	-0.003	0.090***	0.068**	0.023**	0.085**	0.033	0.052***
	(0.026)	(0.026)	(0.007)	(0.024)	(0.024)	(0.006)	(0.025)	(0.025)	(0.008)	(0.026)	(0.027)	(0.009)	(0.036)	(0.036)	(0.014)
Education	No	Yes	-0.002	No	Yes	-0.004	No	Yes	-0.001	No	Yes	0.003	No	Yes	0.001
			(0.003)			(0.003)			(0.003)			(0.004)			(0.004)
Avg GPA	No	Yes	0.002	No	Yes	-0.001	No	Yes	-0.003	No	Yes	0.008*	No	Yes	0.010*
			(0.003)			(0.003)			(0.003)			(0.004)			(0.005)
SES	No	Yes	-0.007	No	Yes	-0.006	No	Yes	-0.008	No	Yes	0.003	No	Yes	0.016**
			(0.005)			(0.005)			(0.005)			(0.006)			(0.008)
Gender	No	Yes	-0.001	No	Yes	-0.001	No	Yes	-0.002	No	Yes	-0.001	No	Yes	0.003
			(0.001)			(0.001)			(0.002)			(0.001)			(0.003)
Network Difference	No	Yes	0.000	No	Yes	0.005*	No	Yes	0.010**	No	Yes	0.009**	No	Yes	0.023**
			(0.002)			(0.003)			(0.004)			(0.005)			(0.010)

- Note:

3

E-learning Engagement Gap During School Closures: Differences by Academic Performance

Joint with Josep Amer-Mestre and Marta C. Lopes

Abstract We study the impact of COVID-19 school closures on differences in online learning usage by regional academic performance. Using data from Google Trends in Italy, we find that during the first lockdown, regions with a previously lower academic performance increased their searches for e-learning tools more than higher-performing regions. Analysing school administrative and survey data before the pandemic, we find that both teachers and students in lower performing regions were using no less e-learning tools than higher performing ones. These two findings suggest that the COVID-19 shock widened the e-learning usage gap between academically lower and higher-performing regions. Exploiting the regional variation in school closure mandates during the 2020-2021 academic year, we report that the patterns detected after the first lockdown were no longer present. Regions with different previous academic performance had the same response in terms of online learning usage when faced with stricter school closures.

3.1 Introduction

Closing schools has been one of the primary measures of governments worldwide to prevent the spread of the COVID-19 virus. As a result, teachers and students were forced to an unprecedented sudden transition from face-to-face to online schooling. Empirical evidence indicates that this pandemic brought short-term average learning losses for students (see e.g. Contini et al., 2022; Maldonado and De Witte, 2021), and disproportionate learning losses for students from lower socioeconomic backgrounds (Engzell et al., 2021). In this context, it is important to understand how different areas of a country responded to the shock, and whether existing regional disparities have been widened. In particular, the extent to which e-learning resources – which have required substantial investments by governments, schools and households – have been used across the different regions of the country during the mandatory school closures is of particular relevance.

In this paper, we study the differential response in online learning usage by regions with different pre-pandemic academic performance during school closures. Most of the literature studying the link between e-learning and academic performance has focused on the possible adverse effects on student outcomes.[1] However, little is known about how students in regions with different academic performances engage with the available e-learning tools when their in-presence class time is reduced (Figlio et al., 2013; Joyce et al., 2015).

We analyse the heterogeneity in e-learning engagement during two periods of school closures. We measure the engagement with online learning resources using real-time data via Google Trends for Italian regions and analyse two distinct periods. One from September 2016 to June 2020, which includes the period in which a nationwide school closure was implemented. A second one from November 2020 to June 2021, in which school lessons were carried out either in-person or online intermittently depending on the local spread of the virus. To measure academic performance, we use pre-pandemic average standardised test scores in reading and

[1]The following papers have found none to negative average effects of e-learning tools on academic performance: Brown and Liedholm (2002), Fairlie and Robinson (2013), Figlio et al. (2013), Joyce et al. (2015), Beuermann et al. (2015), Bando et al. (2017), Cristia et al. (2017), and Lu and Song (2020).

mathematics administered by the National Institute for the Evaluation of the Education and Training System in Italy (INVALSI).[2]

We document four main findings. First, Google search data on selected popular e-learning platforms show a vast surge in their searches right after the nationwide school closure was introduced. We estimate that during the more than three months that schools remained closed there was an average increase of 142 percent in searches of such type of platforms, compared to the pre-pandemic period. This big and yet sustained increase in the Google searches for e-learning platforms is explained by the unprecedented transition that the schooling system had to endure in the first months of the COVID-19 pandemic.

Second, estimating a difference-in-differences specification we find that after schools were closed, regions with lower academic performance experienced a 19 percent *higher* increase in their search for online learning resources compared to higher-achieving regions. This is the opposite result found in the U.S., where it was more affluent areas, with better internet access, and fewer rural schools that saw substantially larger increases in internet search intensity of online learning resource (Bacher-Hicks et al., 2021).

Third, we report that regions with higher average academic performance did not have a higher engagement in online learning in pre-COVID-19 times. Using PISA (OECD Programme for International Student Assessment) and INVALSI surveys, we show a statistically significant negative association between academic performance and the use of online learning resources by students outside school and by teachers in-class at a regional level. Thus, we argue that Italian regions with a higher academic performance did not face the COVID-19 outbreak with a greater familiarity in the use of online learning resources.

Finally, the analysis of the 2020-2021 academic year allows us firstly to ascertain the accuracy of the Google Trends data as a valid proxy for measuring changes in e-learning platform usage. Secondly, to show that previous academic performance was no longer a relevant factor determining differences in e-learning platform usage in the new academic year.

All these results, taken together, suggest that the first months of the pandemic contributed to widening the gap in the usage of online learning resources between academically high and low performing regions in Italy. Not only did lower-performing regions have higher levels of

[2]Established in 1997, among other tasks, INVALSI is entrusted with administering periodic tests to evaluate students' academic achievement at different levels of education.

online learning usage before the pandemic, but during the pandemic, they also increased the search for online learning resources more than their counterparts. However, we find that during the subsequent academic year, regions with different pre-pandemic academic performances did not react differently to localised school closure mandates in terms of online learning usage.

Bacher-Hicks et al. (2021) conclude that the usage of e-learning platforms will be one of the main channels through which the COVID-19 pandemic will likely widen socioeconomic gaps in the U.S. Our findings however, do not point in that direction. Instead, we show that the increase in the short-term educational gaps across academically high and low achieving regions in Italy caused by the pandemic, would not be attributed to a lower engagement with online learning resources by lower-achieving regions.

We reconcile the opposite findings of our study and those reported in Bacher-Hicks et al. (2021) by arguing that both in Italy and the U.S., it was the areas with higher preexisting levels in e-learning usage that saw a higher immediate growth in their search rates after the school closures. Thus, we claim that the heterogeneity in the transition cost faced by students and teachers is the main mechanism behind this result. In other words, we argue that in both countries, areas with higher rates of learning usage during pre-pandemic years were simply better equipped to navigate the transition from in-person to online schooling.

This study focuses on Italy, which is an interesting country to study for at least three reasons. First, due to the rapid spread of the virus and its virulence – which resulted in around 130,000 deaths as of September 2021 – it was the first country to close schools outside Asia, and the country with one of the longest school closure period. Schools remained closed nationwide from March 4 2020 until mid-September 2020. Starting in Fall 2020 and throughout the entire 2021 academic year, the Italian Government used a regionalised system to control the virus spread. Depending on regional outbreaks, each Italian region was assigned a different colour in a four-colour category system. Each of them corresponded to a different set of measures, including school closing mandates. We answer our research question using two different analyses, one for each institutional setting.

Second, Italy is a country that presents substantial regional differences in school quality and academic performance (Agasisti and Vittadini, 2012; Argentin and Triventi, 2015), with a pronounced North-South divide. In a country where achievement gaps across regions are a concern, studying how regions with different academic performances engaged in online learning during the pandemic is relevant. Third and finally, the Italian setting allows the comparison between the engagement of regions on online learning using platforms that the Government promoted at a national level. Right after the announcement of school closures, the centralised school management in the country put forward a website (*didattica a distanza*) to support schools in implementing online learning methods.

Our work adds to the literature on the effect of COVID-19-induced school closures on online learning engagement by focusing on the differential impact across regions by their academic performance. The learning time gap between low and high academic achievers has been studied using time-use survey data in Germany. Grewenig et al. (2021) show that the reduction of daily time spent learning was significantly larger for low achievers than for high achievers, while they do not find differences by students' socioeconomic status.[3]

We also contribute to the literature that exploits Google Trends data. We show that these data can provide useful, reliable, and real-time information for education-related choices not only in the U.S., but also in smaller countries, with lower initial searches on the Google search tool, such as Italy. In this regard, due to the sampling feature of Google Trends, we call attention to the need to download several samples in settings such as ours. Using nationally representative high-frequency data, our paper documents that lower-achieving regions in Italy were no less engaged in online learning during the lockdown – as measured by their searches for e-learning platforms – than high achieving regions.

The results of this paper can help inform future policy responses in education. If the performance gaps widen as a result of the pandemic, the evidence in this paper calls for a greater involvement of the Government than just providing families with access to these platforms in periods when schools are forced to close.[4] If this were the case, our paper is consistent with a subtler channel: for example, lower-achieving regions doing a less efficient use of online resources, where more searches for online learning resources do not translate into better grades.

[3] Andrew et al. (2020) show that the gap in the time used for learning between primary school students from high and low socioeconomic status increased in England.

[4] For example, Carlana and La Ferrara (2021) find that an intervention giving free, individual, online tutoring to disadvantaged students in Italy substantially increased students' academic performance. Angrist et al. (2020) show that SMS and phone calls to parents minimise learning loss when school close.

The remainder of the paper is organised as follows. Section 3.2 explains the necessary institutional background for our analysis and Section 3.3 describes the main data used for it. Section 3.4 presents the empirical strategy while Section 3.5 shows the findings for the impact of the first nationwide school closure on the change in online learning engagement. Section 3.6 provides descriptive evidence on the use of e-learning by regions before the COVID-19 outbreak. Section 3.7 shows additional results for the 2020-2021 academic year, when school closures were imposed depending on the local spread of the virus. Finally, Section 3.8 concludes.

3.2 Institutional Background

Italy was the first European country hit by the COVID-19 in 2020. The first case of the virus in the country was confirmed by January 31st, but both the intensity and speed of new cases were unequal across the country, leading to a highly regionalised impact, as reported by Giuliani et al. (2020). By February 23rd, the first schools started closing in the two most affected regions, Lombardy and Veneto (*zona rossa*) as well as in two neighbouring regions, Piedmont and Emilia-Romagna. On March 4th, *all* schools and universities across the country closed.[5]

Schools remained closed until the end of the academic 2019/2020 year. The starting date of the 2020-2021 school year differed across some of the Italian regions, with the majority of them starting on September 14th, 2021, and each following their own discretion on school closure mandates. The next meaningful legislative change that affected the development of schooling activity was enacted by the November 3rd, 2021 decree. The new decree established a new method to classify each region into three different categories according to its epidemiological risk – yellow, orange and red.[6]

These new measures imposed online learning only to grade 9 students and above in the two lowest risk zones and extended it to grade 7 students and above for the red zone. After the

[5]Five days later, on March 9th, the president declared a national lockdown. On March 11th, all commercial activity except for supermarkets and pharmacies were closed, and on March 21st, the Italian Government closed all non-essential businesses and industries and restricted the movement of people.

[6]Under each category, the Government implemented different measures to contain the spread of COVID-19. These measures mostly regulated social gatherings and events, and the ability to move across cities and regions. Thresholds in the value of specific epidemiological indicators measured at the regional level, such as relative COVID-19 active cases and the share of occupied beds in intensive care units, determined the changes across colour zones.

Christmas holidays, grade 9 students and above were allowed to go back to in-person schooling in yellow and orange zones. However, the number of students allowed in class was capped from 50 to 75 percent of the classroom's usual capacity. This implied that nine graders and older students were organised in a bi-weekly rotation scheme between in-person and e-learning during yellow and orange zones. Table C.1 in the appendix summarises all the online learning mandates and their changes in the 2020-2021 school year.[7]

Together with the measures restricting mobility, at the beginning of the COVID-19 outbreak, the Italian Ministry of Education put specific measures in place to help teachers, students and families transit from face-to-face to e-learning. At the end of February 2020, the Minister of Education announced on the radio the program *Didattica a Distanza* (distance learning, in English). On March 4th, when *all* schools closed in the country and e-learning became mandatory, the Ministry of Education's website made available a new tab with dedicated training webinars and information on different platforms that were constantly updated. The website promoted three platforms: G Suite, provided by Google, (which includes Google Classroom and Google Meetings), Microsoft Office 365 provided by Microsoft, and WeSchool, provided by the Italian main communication company. While all these platforms already existed before the pandemic, their usage was scarce relative to the high popularity that they gained as a result of the COVID-19 outbreak.[8]

3.3 Data

This Section introduces the data sources that are used in our analysis. Namely, Google Trends, INVALSI, and PISA data, as well as additional data used to control for the COVID-19 spread and other relevant regional level data.

[7]In January 2021, a new lower colour category was introduced, “white", where most of the measures present in the yellow category would not be in place. For schooling activity, however, this new white zone imposed the same measures as those present in its subsequent higher category, yellow.

[8]Based on the data collected by Carlana and La Ferrara (2021) on 427 teachers in 76 middle schools all over Italy, by the month of June 2021 more than 96 percent of the teachers were providing synchronous online classes, and around 85 percent of the teachers provided some asynchronous videos additionally−usually no more than one hour per week. Right after the launch of the website, on March 26th, the Italian Ministry of Education passed the Ministerial Decree n.187, which allocated resources as follows: 1) 70 million euro to buy IT devices, such as tablets or computers, to lend temporarily to students in need, as well as to help these students improve their internet connection; 2) 10 million to allow schools to equip themselves with platforms and digital tools useful for distance learning and; c) 5 million euro to train teachers on methodologies and techniques for distance teaching. Due to bureaucracy delays, however, the help did not arrive to all in need.

Google Trends

We rely on Google Trends data to measure the engagement with online learning platforms in each of the Italian regions during COVID-19. Google Trends calculates the fraction of Google searches that are devoted to a given term relative to the overall Google searches within a given geographic area and time period. This ensures that places with the most search volume are not necessarily always ranked highest. Importantly, Google does not provide the actual fraction of searches but a scaled index ranging from 0 to 100. It assigns a value of 100 to the point in time and geographic area with the highest fraction value. We detail this and other particularities of Google Trends data in Appendix B.

A relevant part of our study is to choose platforms that are exclusively designed for e-learning, to avoid confounding between work-from-home and e-learning. For example, while Google Drive can be used by teachers to upload study material, it is also a commonly used cloud storage application by firms. Thus, in our data its increase in popularity during the pandemic would be attributed to a compound effect of the increase in work-from-home and e-learning. Taking this into consideration, we restrict our keyword list to 5 different platforms exclusively dedicated to e-learning: *Studenti.it*, *Scuola.net*, *Edmodo*, *Google Classroom* and *WeSchool*. It is important to note that the first two are fundamentally different from the last three:

Studenti.it is an Italian website for studying support, managed by the Italian schooling books publisher Mondalori Media S.p.A.. It is one of the most visited websites in Italy by high school students, university students and young people looking for training and employment. The website is constantly updated, and it provides students with the subjects of study of the current school year, study material to prepare for the exams, as well as plenty of practical information, including news from the Ministry of Education.

Scuola.net is a project of La Fabbrica. La Fabbrica is a training institution for teaching staff of the Italian school accredited by the Ministry of Education. It is a website dedicated to teachers of various school grades. A platform where they can participate in free educational initiatives and access solutions for digital teaching.

While these first two are websites where students and teachers can get informed about school and teaching related issues, *Edmodo, Google Classroom and WeSchool* are e-learning platforms themselves.[9] The three of them provide similar services, including allowing teachers to set assignments, have work submitted by students, mark, return graded essays, and distribute quizzes and surveys. In a time when all schools were suddenly forced to switch to online teaching, one would expect the use of the 3 e-learning platforms to experience the most dramatic increase – compared to the other two websites.[10]

It is also worth mentioning that because of low search intensity Google Trends could not provide us with the search information on terms related to *Studenti.it* and *Scuola.net* in the two least populated Italian regions. Therefore, the time series referring to this two e-learning platforms contain slightly less observations than the those referring to the other three platforms.

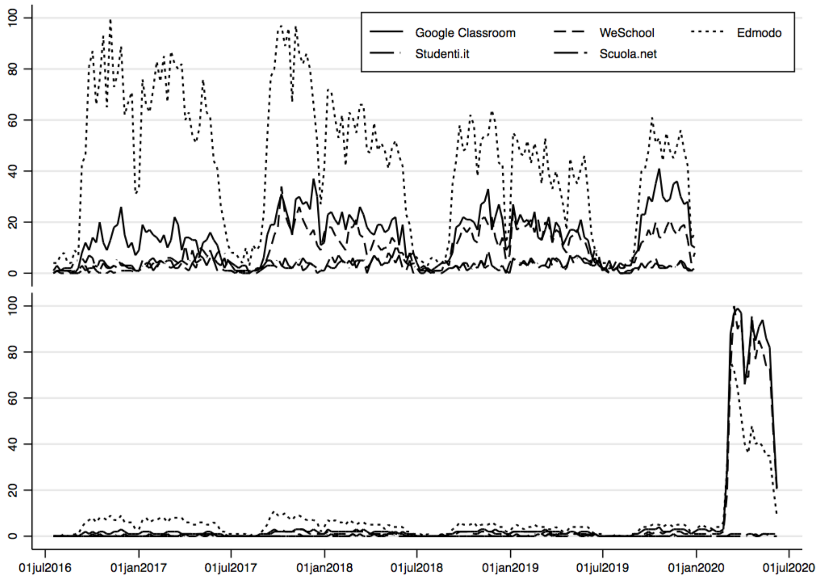
We now show the raw Google Trends data for two purposes: First, to show the features of Google Trends' data and justify our choice regarding how to download our data set. Second, to validate the quality of the data. Each of the two graphs in Figure 3.1 shows 5 series over time, one for each of the 5 keywords described above. The two graphs in the figure correspond to two *different* downloads that differ only on the selected time window. The bottom panel contains the series downloaded for both the before-and-after the pandemic period (from June 2016 to June 2020), while the top panel plots the same data from June 2016 to December 2019. Given that we have exactly 5 keywords per graph, we downloaded the data set corresponding to each graph *all at once*. Thus, the values of this series are comparable *across* – not only within – the series.

The bottom graph in Figure 3.1 shows clear evidence of the dramatic increase of the online search for the three e-learning platforms in Italy right after the COVID-19 outbreak. This increase was lead by Google Classroom, which reached the highest value across all the keyword series on the week of 22-28th of March 2020, thus getting the value 100 in the graph. That same week, WeSchool was searched 91% and Edmodo was searched 60% as much as Google Classroom. Studenti.it and Scuola.net show an almost constant search behaviour over

[9]The website *didattica a distanza*, created by the Italian Ministry of Education as a way to help teachers and students to have a smoother transition into e-learning promoted three different platforms: G Suite, provided by Google (which includes Google Classroom and Google Meetings), Microsoft Office 365 (which includes Word, PowerPoint, Excel, Outlook and Teams), provided by Microsoft, and WeSchool, provided by the Italian main communication company.

[10]In fact, *Google Classroom* and *WeSchool* feature as the third and fourth most searched of all words in the list of ten trending words of Italy during the year 2020, only after *Coronavirus* and *Elezione USA* (USA elections) keywords, which take the first and second places respectively.

Figure 3.1 Google Searches in Italy for 5 selected keywords



Notes: This figure shows the data downloaded from Google Trends for the keywords Google Classroom, WeSchool, Edmodo, Studenti.it and Scuola.net setting the country of Italy as the geographic area. We download two bundles of 5 series each. The first bundle - top graph- contains series spanning from September 2016 to December 2019. The second bundle - bottom graph- spans from September 2016 to June 2020. Given that the series are downloaded in bundles, the series in each graph are comparable within and across across themselves.

the entire time window *relative* to the other three platforms. In that same week, they were each searched 1% as much as Google Classroom.

Figure 3.1 helps to visualize the nature of Google Trends' data, especially when using its comparison feature by downloading the series in bundles. The top panel shows that when downloading the exact same bundle of series excluding the post-pandemic period, the series' variability increases. Thus, we consider that this figure justifies our choice on how to download our data set.

We also use the top graph in Figure 3.1 as supporting evidence that validates the use of Google Trends data to understand the engagement of Italian students with online learning over time. The figure clearly shows that the index of search intensity follows the teaching activity periods along the academic year. The series experience a significant fall during the summer break and fall, to a lesser extent, during Christmas break and Easter holidays. While the level is highest for *Edmodo*, showing that it was the most searched e-learning platform in Italy before the pandemic, Google Classroom and WeSchool followed the same pattern.

Finally, as a further check on the validity of Google Trends' data, Figure C.1 shows that Google Trends is a good predictor of the sudden increase in the number of active Gmail users in Italy in the spring of 2020. We believe that together with Figure 3.1, this is convincing evidence of the validity of Google Trends' data as a measure of engagement in online learning in Italy.

INVALSI

To measure academic performance at the regional level, we use data collected by INVALSI, the National Institute for the Evaluation of the Education and Training System. It organizes yearly standardised tests to assess students' performance at primary school (2$^{\text{nd}}$ and 5$^{\text{th}}$ grades), at lower secondary school (8$^{\text{th}}$ grade), and at higher secondary school (10$^{\text{th}}$ and 13$^{\text{th}}$ grades).

For the purpose of this paper, we focus on students evaluated in the 10$^{\text{th}}$ grade, i.e. higher secondary education. First of all, as students go up on the education system, many of them have extra motivation to study to get access to university, for which there are national entry exams. Second, we give preference to the 10$^{\text{th}}$ rather than the 13$^{\text{th}}$ grade, as these are the students about to complete mandatory education.

In the 10$^{\text{th}}$ grade, students are administered two tests, one on the subjects of reading and one on mathematics, by an external examiner. In Table 3.1 we present the regional rankings of the 2018-2019 academic year.[11] We observe the classic North-South divide for both reading and mathematics. Table 3.1 shows evidence that all regions below the median of both tests are located below Emilia-Romagna. While the ranking position of each individual region is not the same in reading and mathematics, the bundle of regions that lie above the median is the same for both subjects. In all our analysis we use the regional average INVALSI scores for the reading exam.

COVID-19 and Other Data

We now describe the three control variables that we employ in our empirical strategy and their data sources. First, we control for the total number of COVID-19 cases reported daily for each

[11] INVALSI grades are reported according to the WLE (Weighted likelihood estimates) of individual parameters of the Rasch model (Rasch, 1993) where 200 matches the national average.

Table 3.1 Regional Average Scores in INVALSI for 10^{th} Graders

Region	Average reading	Ranking reading	Average math	Ranking math
Valle d'Aosta	218	1	212	6
Lombardia	217	2	217	4
Trento	217	3	224	1
Veneto	216	4	220	2
Friuli-Venezia Giulia	213	5	218	3
Emilia-Romagna	211	6	214	5
Piemonte	210	7	211	8
Marche	210	8	212	7
Liguria	206	9	207	9
Bolzano	206	10	206	11
Umbria	206	11	207	10
Lazio	205	12	200	14
Abruzzo	204	13	203	13
Toscana	203	14	205	12
Molise	199	15	198	15
Basilicata	196	16	193	17
Puglia	196	17	194	16
Campania	192	18	188	18
Sicilia	192	19	184	19
Calabria	189	20	184	20
Sardegna	187	21	182	21

Notes: This table reports INVALSI regional average scores in the 2018-2019 school year. Data obtained from the 2019 INVALSI report.

region, provided by the Ministry of Health's website. Given that COVID-19 first and more severely hit the North of the country, we condition on the number of confirmed COVID-19 to account for different trends in the virus spread that would induce different searches in e-learning platforms. Note that all the regions closed all their schools at a similar time, less than 15 days apart, as explained in Section 3.2.

Second, we control for the regional share of households with internet access in 2019, obtained from the National Statistics Institute (ISTAT), and collected by the Annual Questionnaire of Multiscopes for households in Italy. Although virtually all Italian households live in areas covered by broadband internet – in 2017 the European Commission estimated that 99% of all Italian households lived in areas covered by fixed broadband (Commission, 2017) – not all households use this service. Additionally, as can be seen in Figure C.2, all territories have access to similar levels of average download internet speed levels.

Finally, we include a northern dummy which follows the ISTAT terminology for statistical purposes. This dummy takes the value one for Emilia-Romagna, Friuli-Venezia Giulia, Lombardy, Piedmont, Trentino-Alto Adige, Valle d'Aosta, and Veneto. Italy's North-South divide in terms of cultural, socioeconomic and labor market characteristics is well documented. Thus, this dummy accounts for the North-South differences in all these characteristics, which may in turn drive differences on academic performance and e-learning usage.

3.4 Empirical Strategy

This section lies down the empirical strategy used to analyse the impact that the 2020 school closures had on the change of e-learning platform usage by level of academic performance across regions. Given hat the first case of the COVID-19 virus in Italy was confirmed on January 31st 2020, regional school closures were implemented as follows: the regions of Piemonte, Emilia-Romagna, Lombardy and Veneto closed on February 23rd 2020, Marche and the province of Trento on February 24th, Liguria on March 1st, and on March 4th all the remaining regions closed their schools. Soon after each closure, teaching was moved to online platforms and schools across the country remained closed until the end of the 2019/2020 academic year.

We estimate the following model to study whether there were regional differences on the change in search intensity of e-learning platforms after the school closures by their academic performance:

$$\begin{aligned}\ln(G.T.Index_{j,r,w}) = {} & \alpha_0 + \alpha_1 1AfterSchoolClosure_{r,w} + \beta_2 INVALSIScore_r + \\ & + \beta_3 1AfterSchoolClosure_{r,w} \times INVALSIScore_r + \\ & + \gamma \ln(TotalCases_{r,w}) + X'\delta + \lambda_j + \phi_w + \epsilon_{j,r,w} \qquad (3.1)\end{aligned}$$

$\ln(G.T.Index_{j,r,w})$ is the log of Google Trends index for term *j* in region *r* in week *w*. Note that because the index includes zeros we shift it by one unit so that the dependent variable is defined for all weeks in our time window. $1AfterSchool\ Closure_w$ is an indicator variable that takes value 1 after the week schools closed in region r and 0 before. $INVALSIScore_r$ represents the average score obtained in the 2019 INVALSI test for reading in region r.

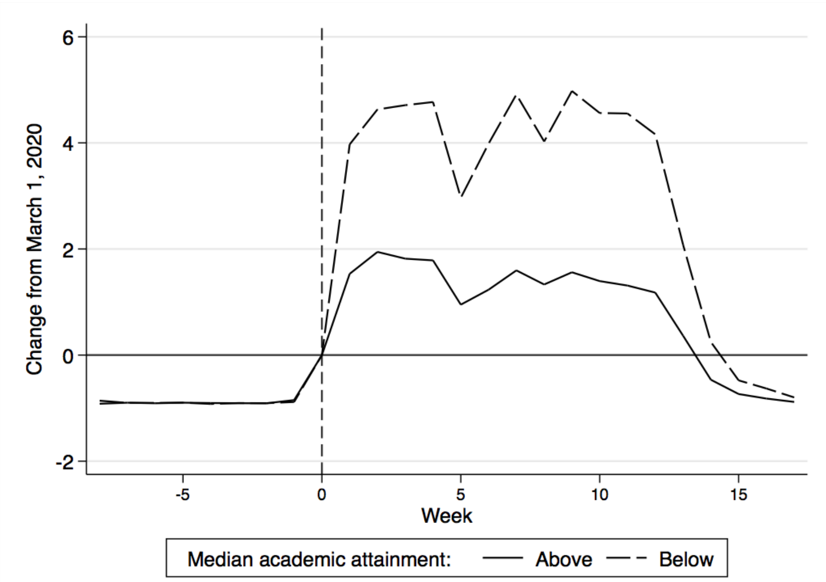
This variable has been standardised – i.e. demeaned and divided by its standard deviations. $\ln(TotalCases_{rw})$ is the total number of notified COVID-19 cases in region *r* in week *w*, to capture the potential increase in the need to use more e-learning rather than alternative in-person resources. X is a matrix of (time-invariant) regional characteristics, which includes: the share of households with internet access at home, to approximate internet usage and the total amount of terms searched in that region; and a dummy for whether the region is in the North of the country, to capture invariant regional characteristics of that part of the country, as well as the fact that they were firstly hit by the virus. To account for seasonality factors, fixed effects for the academic year and week of the year are introduced in ϕ_w. λ_j are platform fixed effects. Finally, $\epsilon_{j,r,w}$ is the error term. Our coefficient of interest is β_3, and it measures the effect of one standard deviation increase in INVALSI scores on the change of e-learning usage after schools closed relative to the period before school closures. Standard errors are clustered at the region level and bootstraped 1000 times to account for the low number of regions in our case study. All coefficients are weighted by the 2019 population values in each region to obtain nationally representative results, and the time window we use is between June 27th 2016 and June 7th 2020.

3.5 Results

We begin by providing a descriptive visualisation of the difference in online learning usage across regions with different levels of academic performance. Figure C.3 depicts the average search indices for regions above and below the 2019 median INVALSI score within a time period of 25 weeks. It clearly illustrates that while academically high and low performing regions have a similar pattern both before and after school's closure, the increase in search intensity is substantially different, with regions below the median 2019 INVALSI score searching more than those above.

Table **??** in the Appendix reports the results from an event study quantifying the total change in the usage of e-learning platforms due to school closures. The first column shows how, on average, regions increased the search of the e-learning platforms terms by 143%, relative to the period before school closures. Compared to southern regions, northern regions

Figure 3.2 Google Trends Search Index for Google Classroom by Academic Performance



Notes: This figure plots weekly changes of the Google Trends search index for the term *Google Classroom* in two groups of regions relative to March 1, 2020. Search index represented under below (above) the 2019 median INVALSI score contain the population weighted mean of the search index for the regions with a score in reading below (above) the national median. Regional mean scores in reading are extracted from the 2019 INVALSI report corresponding to Grade 10 students. Regional population shares used for the weights correspond to 2019 and are extracted from ISTAT.

had on average a 7% higher increase in their searches of e-learning platforms during the entire period.

To perform a more exhaustive analysis, we estimate the regression equation (3.1) and present the results in Table 3.2. The first column pools all search data across the main five e-learning platforms at the national level, while results for each of them are shown in columns 2 to 6. In the third row of the first column we observe that after the closure of schools regions differed in their changes of e-learning platform searches depending on their previous academic performance in the 2019 INVALSI test. Specifically, we estimate that regions scoring one standard deviation above the average INVALSI score in reading had 19% lower change in their internet searches about e-learning platforms. As expected, regions that reported more COVID-19 cases are associated with higher levels of internet searches of e-learning platforms.

Importantly for the identification strategy, the results show that in the period before the school closures the regional academic performance was not an economic nor statistically significant factor associated with an increase of e-learning platform searches (between 3% and 9%), as indicated by the first coefficient of each row.

Analysing the difference-in-differences results by platform we conclude that the differential effect of academic performance on the change of e-learning platform searches after the school closures was driven by three of the five main platforms used in Italy, namely Google Classroom, WeSchool and Studenti.it. In contrast, previous academic performance was not detected to play a statistically significant role in the changes of searches related to Edmodo and Scoula.net platforms.

We find very similar results when using two alternative definitions of school closures, reported in Table C.3. The first alternative definition uses the March 4th 2020 as the school closing date for all regions. The second one drops all observations between February 15th and March 15th 2020, and uses the latter date as the school closure date.[12]

With a similar data set but for the case of the U.S., Bacher-Hicks et al. (2021) show that economically more developed areas of the country saw substantially larger increases in search intensity for online learning platforms. Bearing in mind that areas of the United States with higher income are also areas with higher average SAT scores (Chetty et al., 2020), our analysis shows that the opposite effect took place in Italy. From their result the authors conclude that the pandemic will widen achievement gaps due to these areas' different engagement with online learning resources during the lockdown. Our findings, however, do not support the idea of differences in e-learning platform usage being a relevant factor when trying to explain the differences in immediate school outcomes.

The reasons why our analysis yields opposite findings to those reported for the U.S. are potentially related to the difference in the pre-pandemic usage of e-learning platforms across these two groups of regions. We analyse this for the case of Italy in the next section and we find that, before the outbreak of the pandemic, Italian regions with lower academic achievement were already using e-learning platforms more than those with higher academic achievement. Based on the pre-pandemic levels reported in their analysis and on the results from previous studies (see for instance Vigdor et al., 2014) we know that this was not the case in the U.S. There, students in high academically performing areas were already using e-learning resources more than those in low performing areas before the outbreak of COVID-19.

[12]Table C.2 in the Appendix reports the same type of analysis for the event study.

By acknowledging that in both countries the areas with higher preexisting usage rates of e-learning resources are the ones experiencing higher immediate increases, we are able to reconcile the results of both studies. Finally, we argue that the main mechanism behind this result is related to the cost of transitioning from in-person to online schooling. Students and teachers in areas with higher preexisting usage rates of e-learning resources were more likely to face a significantly lower transition cost compared to those with lower pre-pandemic usage rates. These lower costs could be related to the fact that the required software and hardware was already available in those regions, as well as to the higher skills and familiarity in using them of their teachers and students.

Table 3.2 Difference-in-Difference Results

	(1) All	(2) GC	(3) WS	(4) Ed	(5) Sc	(6) St
INVALSI Score	0.066	0.096	0.059	0.047	0.097	0.030
	(0.202)	(0.095)	(0.047)	(0.101)	(0.537)	(0.484)
After Regional Schools Closure	0.988***	2.436***	2.466***	2.212***	-1.085*	-1.337***
	(0.192)	(0.105)	(0.136)	(0.160)	(0.560)	(0.450)
After Regional Schools Closure * INVALSI Score	-0.188***	-0.285***	-0.193***	-0.088	-0.132	-0.268***
	(0.039)	(0.037)	(0.033)	(0.057)	(0.094)	(0.100)
North	0.002	0.106	0.124	0.159	-0.267	-0.113
	(0.249)	(0.130)	(0.078)	(0.191)	(0.650)	(0.621)
ln(COVID-19 Cases)	0.118***	0.092***	0.109***	0.006	0.182***	0.226***
	(0.021)	(0.014)	(0.015)	(0.019)	(0.063)	(0.054)
Share of Internet Access	0.009	0.011	-0.007	0.005	0.012	0.023
	(0.027)	(0.014)	(0.007)	(0.019)	(0.074)	(0.068)
Constant	0.684	-0.100	1.054*	1.145	0.931	0.391
	(2.092)	(1.078)	(0.569)	(1.421)	(5.734)	(5.230)
Observations	19,776	4,120	4,120	4,120	3,708	3,708
Platform FEs	Yes	-	-	-	-	-
Academic year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Week of the year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Bootstrap replications	1,000	1,000	1,000	1,000	1,000	1,000

Notes: This table reports the results from estimating equation 3.1 by ordinary least squares during the period June 27th 2016 to June 7th 2020. The dependent variable is the logarithm of the Google Search Index for selected E-learning platforms. *After Schools Closure* takes value 1 when schools closed in each region and 0 before. *INVALSI Score* represents the average score obtained in 2018 in the INVALSI test for Italian language. This variable has been standardised (demeaned and divided by its standard deviations) hence its units are standard deviations. *North* takes value 1 for Emilia-Romagna and all regions above it, and 0 otherwise. *Share of Internet Access* contains the share of households in each region that had internet access in 2019. *ln(COVID-19 Cases)* contains the total number of COVID-19 cases reported in each region and day. GC stands for Google Classroom, WS for WeSchool, Ed for Edmodo, Sc for Scuola.net and St for Studenti.it. All regression coefficients are weighted by each region's population and include fixed effects for each searched platform, academic year and week of year. Bootstrapped standard errors are clustered by region and reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

3.6 Online Learning Engagement before the Pandemic

Google Trends' Index values, allows us to study which set of regions *changed* the search intensity more as a result of the pandemic induced school closures. We just showed that contrary to findings in other studies, during the pandemic in Italy, academically lower performing regions *increased* the engagement with online learning platforms more than academically higher performing regions.

To interpret this result it is important to explore which of two opposite mechanisms, both consistent with our finding, is likely to have prevailed: 1) A catching-up-effect where academically lower-performing regions faced the COVID-19 outbreak with a lower *level* of engagement, and thus, had a bigger room for improvement; or 2) a gap-widening effect where academically lower-performing regions already had a higher engagement, and during the pandemic widened this gap even more.

For this, we need to compare the *levels* of engagement with e-learning before the pandemic across regions with different academic performances. Unfortunately, our Google Trends' Index data set does not allow to do so, and thus, we have to rely on other data sources.[13]

To analyse the relationship between academic performance and the $level$ of online learning usage before the pandemic, we would ideally like to have the number of users and accesses, by region, to each of the three e-learning platforms and two websites that we use in our main analysis. Unfortunately, the data is not available. Thus, we have to rely on other data sources, and we use PISA and INVALSI as they are the two most complete surveys related to education in Italy. Taken together, they present a piece of consistent descriptive evidence that academically higher-performing regions were *not* using online learning more *before* the pandemic.

[13]As explained in Section 3.3, the value of the index for a given term in each of the series – corresponding to each of the regions – is a value relative to each series' own peak i.e if Lombardy takes the value of 70 and Campania takes the value of 50 on the index on a given date for a given term, it means that in that particular date, that term was searched 70% as much as in its most searched day in Lombardy and 50% as much as in its most searched day in Campania. We still do not know whether in that day and for that term, Lombardy had a higher search intensity than Campania or the opposite was true.

3.6.1 Use of e-learning tools before the pandemic by students

PISA (Programme for International Student Assessment) is an international standardised survey to 15-year-old students that comprises of a cognitive test on reading, mathematics and science, and complementary questionnaires to assess students' attitudes and motivations. Two surveys, the ICT Familiarity Questionnaire and the Educational Career Questionnaire are relevant to us. While the questionnaires have a very rich set of questions, the caveat of PISA is that not all the regions participate in every wave. We use PISA 2015 because it includes the better-suited regions for this study, Lombardy and Campania.[14] The two regions are among the most populated regions and have already been used as representative cases of the north-south divide in Italy in other studies (Acconcia and Graziano, 2017). We provide results comparing the two of them, where we use Lombardy as an example of the academically higher-performing regions of the North and Campania as an example of the lower-performing regions of the South.

From the various questions available, we focus on three that assess the ICT usage and availability outside school, as the availability and usage at school will be discussed in the data reported from teachers to INVALSI in the next subsection. Panel A in Table 3.3 reports differences in the usage of ICT resources for schoolwork and Panel B to attend additional instructions which are not part of students' mandatory school schedule.

Panel A shows clear evidence that in the year 2015 students in Campania were already using e-learning technologies for schoolwork outside school more than students in Lombardy. Students in Campania were 10.4% more likely to use the internet for schoolwork, 45.1% more likely to use the internet to follow-up school lessons, 23.3% more likely to do their homework using a computer and 56.4% more likely to do them using a mobile phone. As reported in the third column of Table 3.3, all these differences are statistically significant at a 1% level. Panel B shows that in 2015 students in Campania were also more likely to use ICT in their additional

[14]PISA 2015 provides data for Bolzano, Campania, Lombardy and Trento, while PISA 2018 provides data for Bolzano, Toscana, Sardegna and Trento. Note that both Bolzano and Trento (which form Trentino-Alto Adige) have a considerably lower share of publicly managed schools and therefore might be using e-learning differently than schools managed by the State. Excluding these two regions, PISA 2018 does not include any other region from the "above median performance" group we consider in our main analysis. Therefore, PISA 2015 is best suited for our analysis.

Table 3.3 ICT Usage in Selected Regions

Variable: Proportion of students	Campania (1)	Lombardy (2)	Difference (3)	Italy (4)
Panel A				
Outside school, at least once a week				
- for schoolwork	0.626	0.567	0.060***	0.591
	(0.013)	(0.013)	[0.001]	(0.009)
- to follow up school lessons	0.602	0.415	0.187***	0.504
	(0.014)	(0.013)	[0.000]	(0.009)
- for doing homework on computer	0.423	0.343	0.080***	0.362
	(0.014)	(0.012)	[0.000]	(0.009)
- for doing homework on mobile	0.416	0.266	0.150***	0.322
	(0.014)	(0.012)	[0.000]	(0.009)
Panel B				
Additional Math Instructions				
- Internet tutoring by a person or app	0.235	0.162	0.073***	0.185
	(0.017)	(0.016)	[0.002]	(0.011)
- Video recorded	0.168	0.069	0.099***	0.111
	(0.015)	(0.011)	[0.000]	(0.009)
Additional Italian Instructions				
- Internet tutoring by a person or app	0.275	0.226	0.049	0.263
	(0.020)	(0.023)	[0.112]	(0.016)
- Video recorded	0.155	0.103	0.052**	0.130
	(0.017)	(0.016)	[0.027]	(0.012)

Notes: The data reported in Panels A and B come from PISA 2015 ICT Familiarity Questionnaire and Educational Career Questionnaire respectively. Columns 1,2, and 4 report the proportion of students that answered positively to each of the metrics. Standard errors are reported in parenthesis. Column 3 reports the difference between Campania and Lombardy. The stars ,***,**,*, in this column indicate whether the difference is statistically significant at 1%,5%, and 10%, respectively. The p-values associated with the differences tests are reported in square brackets. All averages are weighted by the PISA final trimmed non-response adjusted student weights.

instructions (not part of the student's mandatory school schedule) in both mathematics and reading.[15]

3.6.2 Use of e-learning tools before the pandemic by teachers

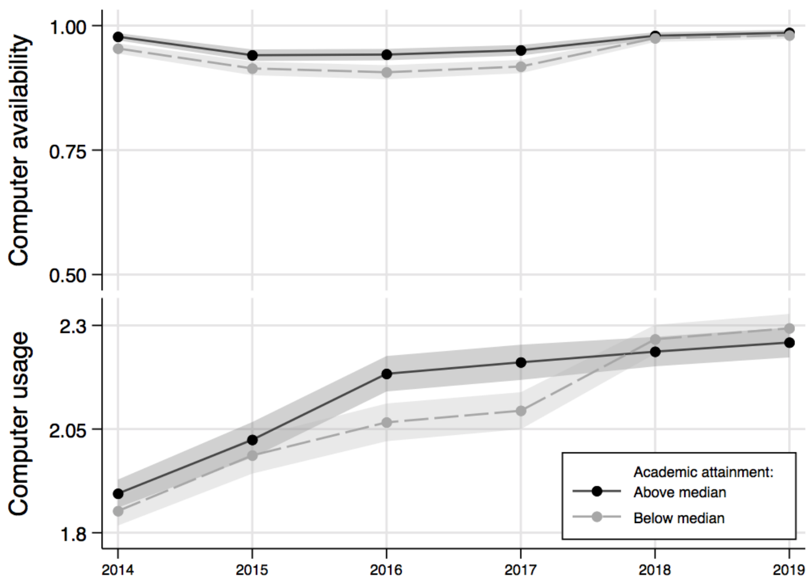
Together with the tests described in Section 3.3, INVALSI carries out surveys to students, teachers and school principals. The advantage of INVALSI over PISA is that the former includes a representative random sample of schools for every region in Italy. We begin

[15]Despite Campania being a much poorer region than Lombardy, one could wonder if results are driven by higher access to ICT, by students in Campania. We use the ICT Familiarity Questionnaire, which asks about device availability at home, and find that this is not the case.

by showing computer availability in schools by regions above and below the median 2019 INVALSI score.

Figure 3.3 plots the proportion of Italian language and mathematics teachers reporting having access to a computer and their usage in class during their lessons in every academic year between 2013 and 2019. The top figure suggests that until the academic year of 2017/2018, in higher academically performing regions a higher proportion of teachers had access to a computer to conduct their lessons. However, by 2018 and onward the two rates converged. The bottom figure shows the same pattern for computer usage and confirms that, in the last two academic years, there has been no difference in computer usage between low and high academic performing regions.

Figure 3.3 Computer Availability and Usage by Teachers in Class by Academic Performance

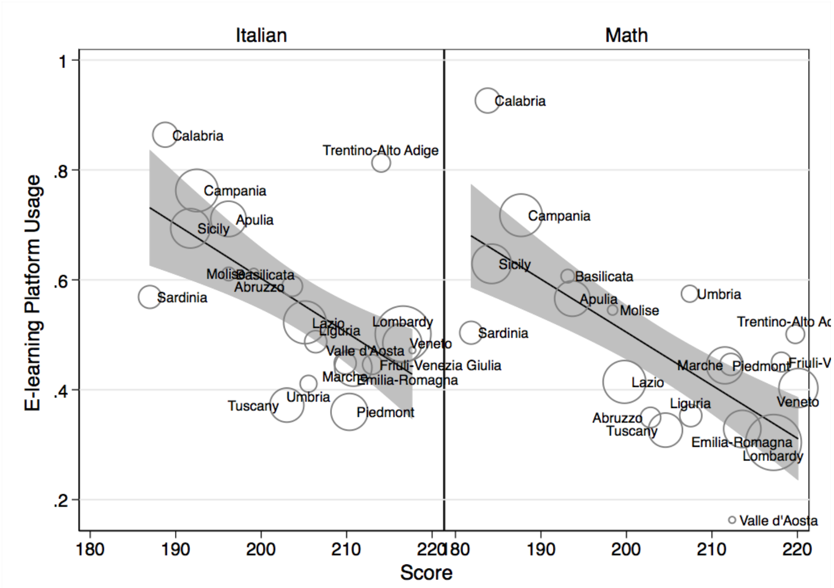


Notes: The figure plots the proportion of Italian language and mathematics teachers reporting having access to a computer, panel (a), and their usage, panel (b), in class during their lessons. Values are taken from a specific responses to question D6a administered by INVALSI to Grade 10 teachers of both subjects from 2013 to 2019, which states: *How much did you use the computer in lessons with the students of your class in the last school year?* Panel (a) plots one minus the share of teachers who responded *No computer present in school*. Panel (b) plots the group average of the following response options: *0 = No computer present in school; 1 = I don't use it; 2 = Occasional use; 3 = Regular use*. Below (above) the median contains the mean of the responses in the regions with a mean score in each subject below (above) the national median, respectively. Regional mean scores in both subjects are extracted from the 2019 INVALSI report corresponding to Grade 10 students.

Unfortunately, the INVALSI surveys to students do not ask about their use of online tools outside school. However, the teachers' questionnaire includes a question of our interest: "Thinking about the didactic activity you carried out this year, please indicate how often you carried out the following activities: use of e-learning platforms.", the response options being 0 =Never or almost never; 1 = Sometimes; 2 = Often; 3 = Always or almost always.

Figure 3.4 plots Grade 10 teachers' reported usage of e-learning platforms when conducting their didactic activity in academic year 2018/2019 in reading and mathematics classes by their INVALSI score on those subjects in that year. Results show that lower academic performance regions are the ones associated with a greater level of e-learning usage in school by teachers.

Figure 3.4 Teachers' E-learning Platform Usage by Students' Academic Performance



Notes: This figures shows the correlation between the reported usage of e-learning platforms by teachers when conducting their didactic activity in each region with the average results for the 2018/2019 INVALSI tests in reading and mathematics at Grade 10. The usage values for e-learning platforms are taken from the responses to question: *Thinking about the didactic activity you carried out this year, please indicate how often you carried out the following activities: e) use of e-learning platforms.* With the following response options: *0 = Never or almost never; 1 = Sometimes; 2 = Often; 3 = Always or almost always.* Sizes of circles correspond to the population share of each region, in 2019. The solid line corresponds to a linear fit weighted by the population share of each region. The shaded area corresponds to a 95% confidence interval of the linear fit.

Summing up, with these two data, we show that lower-performing regions were not using online learning tools at a lower level than higher-performing regions before the pandemic, allowing us to reconcile our results with those found by Bacher-Hicks et al. (2021) in the U.S.

3.7 Additional Results

As explained in section 3.2, from November 6th, 2020 onwards the Italian Government categorised regions in three categories using colours: yellow, orange and red, according to the different levels of the spread of the virus. Under each category, different measures were implemented to contain the spread of COVID-19 during the second wave. The new measures imposed online learning to grade 7 and above in the Red zone, while grade 9 students and above had to also follow their lessons online in the two lowest risk zones, yellow and orange. After the Christmas holidays, grade 9 students and above were allowed to go back to in-person schooling during yellow and orange. However, the number of students allowed in class was capped from 50 to 75 percent of the classroom's usual capacity. This implied that nine graders and older students were organised in a bi-weekly rotation scheme between in-person and e-learning during yellow and orange zones. Table C.1 in the appendix summarises all the online learning mandates and their changes in the 2020-2021 school year. In this section we want to exploit the variation in the change of online learning platforms usage imposed by the new zone classification to analyse whether the pattern discovered after the first nationwide schools closure still persisted.

For this, we use Google Trends' data from the 2020-2021 academic year alone together with daily information on the assigned colour zone for each region. Importantly for our analysis, regions were declared at different moments and with different frequencies into the most restrictive category, the Red zone. Table C.4, in the Appendix, summarises the descriptive statistics of the assignment of the regions to each of the colour zones. The data collection ranges from the beginning of the new zoning system, November 6th, to the end of the 2020-2021 academic year, June 18th.

In this analysis we seek to exploit the variation introduced by categorising regions into different colour zones to analyse two subjects. First, we are interested in testing the accuracy of our Google Trends' measure as a proxy for e-learning usage by exploiting the regional variation in online learning mandates. Second, we want to understand whether the result found for the previous academic year, that is, that students in lower academic performing regions increased their usage of e-learning platforms more, is still present in the new course. To perform each of these two analyses, we pool daily Google Trends data on the three main e-learning platforms used across the different levels of education – Google Classroom, WeSchool and Edmodo – and estimate the two following specifications:

$$
\begin{aligned}
\ln(G.T.Index_{j,r,d}) = {} & \alpha_0 + \alpha_1 1RedZone_{r,d} + \alpha_2 1OrangeZone_{r,d} + \beta_1 INVALSIScore_r \\
& + \gamma_1 \ln(TotalCases_{r,d}) + X'\delta + \lambda_j + \phi_w + \epsilon_{j,r,d}
\end{aligned} \tag{3.2}
$$

$$
\begin{aligned}
\ln(G.T.Index_{j,r,d}) = {} & \alpha_0 + \sum_c \alpha_c 1ZoneC_{r,d} + \beta_2 INVALSIScore_r + \\
& + \sum_c \delta_c 1ZoneC_{r,d} \times INVALSIScore_r + \\
& +_1 \ln(TotalCases_{r,d}) + X'\gamma + \lambda_j + \phi_w + \epsilon_{j,r,d}
\end{aligned} \tag{3.3}
$$

In equation (3.2) we are interested in measuring the correlation of changes of colour zones with changes in e-learning usage, measured by α_1 and α_2. $1RedZone_{r,d}$ and $1OrangeZone_{r,d}$ take value 1 if region r in day d was declared to be in the Red or Orange zones and zero otherwise, respectively. Since weekends and national holidays are removed from our studied sample, the base group of the colour indicator variables aggregate both the Yellow and the White zones. Thus, the base group is expected to contain the periods with low e-learning usage. λ_j are platform fixed effects and ϕ_w week of the year fixed effects. The rest of the variables are defined as explained in equation (3.1). Standard errors are clustered at the region level. We bootstrap the standard errors 1000 times to account for the low number of regions in our case study. All coefficients are weighted by the 2019 population values in each region to obtain nationally representative results.

In (3.3) we test if regions in the same colour zone, and therefore with the same online learning mandate, present a different change in their e-learning usage according to their average academic grade at the 2019 INVALSI test on reading. To do so, we interact the indicator variables associated with each region's colour zone and day ($1ZoneC_{r,d}$ equals one if region r is in colour c, c being either red or orange at day d) with the standardised INVALSI score.

The first column in Table 3.4 reports the estimates of equation (3.2). As expected, compared with periods in which regions are declared Yellow or White, the change in the usage of e-learning increased more as online learning mandates were declared for a higher number of students; that is, when regions turned into orange or red zones, respectively. Also, it is no surprise that the change in search of e-learning resources is larger when changing to Red zone, 34.5 percent, than when doing it to Orange, 9.7 percent. These estimates are statistically significant at the 1 percent level. These first results confirm that our measure of changes in the Google searches of e-learning platforms is a proper proxy for the actual change in the platforms' usage.

Table C.5 reproduces the same type of analysis using data only for one of the three main platforms each time. The largest change in the searches when entering in each new colour zone was coming from WeSchool platform, followed closely by the change in searches for Google Classroom. No statistically significant changes are detected for the changes in searches of Edmondo.

The second column reports the estimates of equation (3.3). We first observe how the estimated change in the search level of the average region in terms of the 2019 INVALSI score when the region is declared to be in the Orange or the Red zones are very similar to those in column 1. The coefficients of the interaction of each colour zone indicator variable with the 2019 INVALSI score are very small and not statistically different from 0. Implying that the regions with different levels of academic achievement did not change their searching behaviour differently during the establishment of each colour category.

We interpret this absence of different changes in e-learning usage between high and low academic performance regions during stricter school closure mandates as suggesting that, after half a year from the COVID-19 outbreak, all regions adapt their online learning behaviour in the same way. That is, the gap between higher and lower e-learning usage regions did not widen.

Table 3.4 Difference-in-Differences Results - Academic Year 2020/2021

	(1)	(2)
INVALSI Score	0.004	-0.005
	(0.134)	(0.136)
Orange Zone	0.097***	0.093***
	(0.033)	(0.034)
Red Zone	0.345***	0.342***
	(0.105)	(0.123)
INVALSI Score x Orange Zone		-0.021
		(0.040)
INVALSI Score x Red Zone		0.065
		(0.100)
North	-0.407*	-0.406*
	(0.212)	(0.211)
ln(COVID-19 Cases)	0.679***	0.679***
	(0.055)	(0.056)
Share of Internet Access	-0.038*	-0.037*
	(0.021)	(0.021)
Constant	-2.672*	-2.735*
	(1.480)	(1.486)
Observations	8,160	8,160
Platform FEs	Yes	Yes
Week of the year FEs	Yes	Yes

Note: This table reports the results from estimating equation (3.2) and (3.3). The sample used contains daily observations from November 6th 2020 to June 7 2021, except for weekends and national holidays. The dependent variable is the logarithm of the Google Search Index for *Google Classroom*. *Red Zone* and *Orange Zone* take value 1 when a region is, respectively, red or orange zone in a certain day and 0 otherwise. *INVALSI Score* contains the regional average score of the 2019 INVALSI test in Italian. *North* takes value 1 for Emilia-Romagna and all regions above, and 0 otherwise. *Share of Internet Usage* contains the share of households in each region that used internet in 2019. *COVID-19 Cases* contains the total number of COVID-19 cases reported in each region and day. Bootstrapped standard errors are clustered by region and reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3.5 Difference-in-Differences Results by Platform - Academic Year 2020/2021

	(1) GC	(2) GC	(3) WS	(4) WS	(5) Ed	(6) Ed
INVALSI Score	0.029	0.036	0.153	0.138	-0.171	-0.188
	(0.110)	(0.104)	(0.143)	(0.155)	(0.349)	(0.349)
Orange	0.093**	0.088**	0.168**	0.169**	0.029	0.021
	(0.037)	(0.036)	(0.071)	(0.076)	(0.086)	(0.094)
Red	0.228***	0.226**	0.522***	0.522***	0.285	0.280
	(0.078)	(0.090)	(0.158)	(0.176)	(0.189)	(0.207)
INVALSI Score x Orange		-0.043		0.021		-0.042
		(0.027)		(0.084)		(0.080)
INVALSI Score x Red		0.017		0.045		0.132
		(0.057)		(0.143)		(0.131)
North	-0.322***	-0.321***	-1.053***	-1.054***	0.155	0.157
	(0.113)	(0.111)	(0.251)	(0.253)	(0.499)	(0.499)
ln(COVID-19 Cases)	0.326***	0.327***	0.810***	0.809***	0.902***	0.901***
	(0.061)	(0.061)	(0.078)	(0.079)	(0.138)	(0.138)
Share of Internet Access	-0.012	-0.011	-0.082***	-0.082***	-0.020	-0.018
	(0.017)	(0.016)	(0.023)	(0.023)	(0.055)	(0.056)
Constant	0.424	0.361	-0.695	-0.695	-7.745*	-7.871*
	(1.375)	(1.351)	(1.775)	(1.800)	(4.098)	(4.123)
Observations	2,720	2,720	2,720	2,720	2,720	2,720
Week of the year FEs	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results from estimating equation (3.2) and (3.3) for each platform. The sample used contains daily observations from November 6th 2020 to June 7 2021, except for weekends and national holidays. The dependent variable is the logarithm of the Google Search Index for *Google Classroom*. *Red Zone* and *Orange Zone* take value 1 when a region is, respectively, red or orange zone in a certain day and 0 otherwise. *INVALSI Score* contains the regional average score of the 2019 INVALSI test in Italian. *North* takes value 1 for Emilia-Romagna and all regions above, and 0 otherwise. *Share of Internet Usage* contains the share of households in each region that used internet in 2019. *COVID-19 Cases* contains the total number of COVID-19 cases reported in each region and day. GC stands for Google Classroom, WS for WeSchool, Ed for Edmodo. Bootstrapped standard errors are clustered by region and reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

3.8 Conclusion

In this paper, we study whether academically high and low performing regions had a different response, in terms of changes in their e-learning usage, to schools' closure mandates imposed by the spread of the COVID-19 epidemic in Italy. We use real-time Google searches to measure the change in the use of several popular e-learning platforms. We divide our analysis into two periods: one spanning from September 2016 to June 2020, which includes the pre-covid time window and the time in which a nationwide school closure was implemented. A second one spanning from November 2020 to June 2021, in which lessons were carried out in-person or online intermittently depending on the local spread of the virus. To measure academic performance, we rely on pre-pandemic average standardised test scores in reading and mathematics administered by INVALSI.

We begin by using real-time Google search data to study the $change$ in the use of online learning tools between regions with different academic performances due to the pandemic. We first document a substantial increase in the usage of e-learning platforms nationwide. Then, using a difference-in-differences specification, we find that regions with lower academic performance *increased* their search for online learning resources *more* after the nationwide school closure was implemented. We further document that previous academic performance was no longer a relevant factor determining changes in e-learning platform usage in the subsequent academic year. We interpret this result as evidence favouring all regions having the same online learning behaviour when faced with stricter school closure mandates during the subsequent academic year.

Several particularities of Google search data limit the policy recommendations that can be extracted from our analysis. Most notably, the fact that the Google Trends data is expressed using a scaled index rather than the actual search levels (or their share over total searches) is indeed its most restricting feature. In particular because it only allows us to speak about the magnitude of the change in Google searches on e-learning resources and not about their usage levels. In our case we rely on alternative data sources to document the level of online learning usage in low and high academically performing regions.

We overcome this limitation by relying on PISA and INVALSI surveys to document the $level$ of online learning usage in regions with different average academic performances before the outbreak of the pandemic. We find that regions with a higher average academic achievement did *not* have a higher engagement in online learning in pre-COVID times. This allows us to argue that Italian regions with higher academic performance did *not* face the COVID-19 outbreak with greater familiarity in the use of online learning resources.

Our results, taken together, suggest that the first months of the pandemic contributed to widening the gap in the use of online learning resources between academically high and low performing regions in Italy. Combining different data, before 2020 and during the subsequent academic year, we have ruled out the channel of a lower engagement in online learning resources by students in lower-achieving regions. The empirical evidence in this paper suggests the need for a greater involvement of governments than just providing households and schools with access to online learning platforms.

References

Abramitzky, R., Boustan, L. P., and Eriksson, K. (2012). Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review*, 102(5):1832–1856.

Acconcia, G. and Graziano, P. R. (2017). *The Youth Guarantee: The Cases of Milan (Lombardy) and Naples (Campania)*, pages 339–348. Palgrave Macmillan UK.

Adda, J., Dustmann, C., and Görlach, J.-S. (2022). The Dynamics of Return Migration, Human Capital Accumulation, and Wage Assimilation. *The Review of Economic Studies*.

Agasisti, T. and Vittadini, G. (2012). Regional economic disparities as determinants of student's achievement in italy. *Research in Applied Economics*, 4(2):33.

Albarrán, P., Carrasco, R., and Ruiz-Castillo, J. (2017). Are migrants more productive than stayers? some evidence from a set of highly productive academic economists. *Economic Inquiry*, 55(3):1308–1323.

Alesina, A., Miano, A., and Stantcheva, S. (2018). Immigration and Redistribution. *NBER Working Paper Series*, 24733:128.

Amior, M. (2019). Education and geographical mobility: the role of the job surplus.

Andrew, A., Cattan, S., Costa Dias, M., Farquharson, C., Kraftman, L., Krutikova, S., Phimister, A., and Sevilla, A. (2020). Inequalities in children's experiences of home learning during the covid-19 lockdown in england. *Fiscal Studies*, 41(3):653–683.

Angrist, N., Bergman, P., and Matsheng, M. (2020). School's out: Experimental evidence on limiting learning loss using "low-tech" in a pandemic. Technical report, National Bureau of Economic Research.

Arcidiacono, P., Hotz, V. J., and Kang, S. (2012). Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics*, 166(1):3–16.

Arcidiacono, P., Hotz, V. J., Maurel, A., and Romano, T. (2020). Ex Ante Returns and Occupational Choice. *Journal of Political Economy*, 128(12):4475–4522.

Arellano-Bover, J. (2020). Career consequences of firm heterogeneity for young workers: First job and firm size.

Arellano-Bover, J. (2022). The effect of labor market conditions at entry on workers' long-term skills. *Review of Economics and Statistics*, 104(5):1028–1045.

Argentin, G. and Triventi, M. (2015). The north-south divide in school grading standards: New evidence from national assessments of the italian student population. *Italian Journal of Sociology of Education*, 7(2).

Attanasio, O. P. and Kaufmann, K. M. (2014). Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender. *Journal of Development Economics*, 109:203–216.

Attanasio, O. P. and Kaufmann, K. M. (2017). Education choices and returns on the labor and marriage markets: Evidence from data on subjective expectations. *Journal of Economic Behavior & Organization*, 140:35–55.

Bacher-Hicks, A., Goodman, J., and Mulhern, C. (2021). Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time. *Journal of Public Economics*, 193:104345.

Baert, S., Cockx, B., and Verhaest, D. (2013). Overeducation at the start of the career: stepping stone or trap? *Labour Economics*, 25:123–140.

Bah, T. L. and Batista, C. (2020). Why do people migrate irregularly? Evidence from a lab-in-the-field experiment in west Africa. *Kellogg Working Paper*, 435:45.

Bando, R., Gallego, F., Gertler, P., and Fonseca, D. R. (2017). Books or laptops? the effect of shifting from printed to digital delivery of educational content on learning. *Economics of Education Review*, 61:162–173.

Barone, G., d'Alessandro, A., and de Blasio, G. (2019). A ticket to ride: Education and migration from lagging areas. *Papers in Regional Science*, 98(5):1893–1902.

Bartik, T. J. (2020). Using place-based jobs policies to help distressed communities. *Journal of Economic Perspectives*, 34(3):99–127.

Bartolucci, C., Villosio, C., and Wagner, M. (2018a). Who Migrates and Why? Evidence from Italian Administrative Data. *Journal of Labor Economics*, 36(2):551–588.

Bartolucci, C., Villosio, C., and Wagner, M. (2018b). Who migrates and why? evidence from italian administrative data. *Journal of Labor Economics*, 36(2):551–588.

Belot, M. V. K. and Hatton, T. J. (2012). Immigrant Selection in the OECD*: *Immigrant selection in the OECD*. *The Scandinavian Journal of Economics*, 114(4):1105–1128.

Bertoli, S., Fernández-Huertas Moraga, J., and Ortega, F. (2013). Crossing the border: Self-selection, earnings and individual migration decisions. *Journal of Development Economics*, 101:75–91.

Beuermann, D. W., Cristia, J., Cueto, S., Malamud, O., and Cruz-Aguayo, Y. (2015). One laptop per child at home: Short-term impacts from a randomized experiment in peru. *American Economic Journal: Applied Economics*, 7(2):53–80.

Blass, A. A., Lach, S., and Manski, C. F. (2010). Using elicited choice probabilities to estimate random utility models: Preferences for electricity reliability. *International Economic Review*, 51(2):421–440.

Boneva, T., Golin, M., and Rauh, C. (2022). Can perceived returns explain enrollment gaps in postgraduate education? *Labour Economics*, 77:101998.

Borjas, G. J., Kauppinen, I., and Poutvaara, P. (2019). Self-selection of Emigrants: Theory and Evidence on Stochastic Dominance in Observable and Unobservable Characteristics. *The Economic Journal*, 129(617):143–171.

Bover, O. (2015). Measuring expectations from household surveys: New results on subjective probabilities of future house prices. *SERIEs*, 6(4):361–405.

Brown, B. W. and Liedholm, C. E. (2002). Can web courses replace the classroom in principles of microeconomics? *American Economic Review*, 92(2):444–448.

Büchel, K., Ehrlich, M. V., Puga, D., and Viladecans-Marsal, E. (2020). Calling from the outside: The role of networks in residential mobility. *Journal of Urban Economics*, 119:103277.

Carlana, M. and La Ferrara, E. (2021). Apart but Connected: Online Tutoring and Student Outcomes during the COVID-19 Pandemic. *HKS Working Paper*, RWP21-001.

Chetty, R., Friedman, J. N., Saez, E., Turner, N., and Yagan, D. (2020). Income Segregation and Intergenerational Mobility Across Colleges in the United States*. *The Quarterly Journal of Economics*, 135(3):1567–1633.

Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics*, 133(3):1163–1228.

Chiquiar, D. and Hanson, G. H. (2005). International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States. *Journal of Political Economy*, 113(2):239–281.

Clemente, J., Larramona, G., and Olmos, L. (2016). Interregional migration and thresholds: evidence from spain. *Spatial Economic Analysis*, 11(3):276–293.

Commission, E. (2017). Europe's digital progress report 2017: Connectivity. page 41.

Contini, D., Di Tommaso, M. L., Muratori, C., Piazzalunga, D., and Schiavon, L. (2022). The COVID-19 Pandemic and School Closure: Learning Loss in Mathematics in Primary Education. *SSRN Electronic Journal*.

Cortés, P., Pan, J., Reuben, E., Pilossoph, L., and Zafar, B. (2022). Gender differences in job search and the earnings gap: Evidence from the field and lab. *The Quarterly Journal of Economics*.

Cristia, J., Ibarrarán, P., Cueto, S., Santiago, A., and Severín, E. (2017). Technology and child development: Evidence from the one laptop per child program. *American Economic Journal: Applied Economics*, 9(3):295–320.

Dahl, G. B. (2002). Mobility and the Return to Education: Testing a Roy Model with Multiple Markets. *Econometrica*, 70(6):2367–2420.

Dahl, M. S. and Sorenson, O. (2010). The migration of technical workers. *Journal of Urban Economics*, page 13.

De la Roca, J. (2017). Selection in initial and return migration: Evidence from moves across Spanish cities. *Journal of Urban Economics*, 100:33–53.

De la Roca, J. D. L. and Puga, D. (2017). Learning by Working in Big Cities. *The Review of Economic Studies*, 84(1):106–142.

Delavande, A. (2008). Pill, pathch, or shot? subjective expectations and birth control choicePILL,. *International Economic Review*, 49(3):999–1042.

Delavande, A. and Zafar, B. (2019). University Choice: The Role of Expected Earnings, Nonpecuniary Outcomes, and Financial Constraints. *journal of political economy*, 127(5):51.

Dominitz, J. and Manski, C. F. (1997). Using Expectations Data to Study Subjective Income Expectations. *Journal of the American Statistical Association*, 92(439):855–867.

Dustmann, C., Fadlon, I., and Weiss, Y. (2011). Return migration, human capital accumulation and the brain drain. *Journal of Development Economics*, 95(1):58–67.

Dustmann, C. and Glitz, A. (2011). Migration and education. In *Handbook of the Economics of Education*, volume 4, pages 327–439. Elsevier.

Dustmann, C. and Görlach, J.-S. (2016). The Economics of Temporary Migrations. *Journal of Economic Literature*, 54(1):98–136.

Dustmann, C. and Weiss, Y. (2007). Return Migration: Theory and Empirical Evidence from the UK. *British Journal of Industrial Relations*, 45(2):236–256.

Engzell, P., Frey, A., and Verhagen, M. D. (2021). Learning loss due to school closures during the covid-19 pandemic. *Proceedings of the National Academy of Sciences*, 118(17).

Fairlie, R. W. and Robinson, J. (2013). Experimental evidence on the effects of home computers on academic achievement among schoolchildren. *American Economic Journal: Applied Economics*, 5(3):211–40.

Fernández-Huertas Moraga, J. (2011). New Evidence on Emigrant Selection. *Review of Economics and Statistics*, 93(1):72–96.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*, volume 2 of *Stanford University Press.*

Figlio, D., Rush, M., and Yin, L. (2013). Is it live or is it internet? experimental estimates of the effects of online instruction on student learning. *Journal of Labor Economics*, 31(4):763–784.

FundaciónBBVA (2018). Movilidad Universitaria. ESENCIALES Fundación BBVA- Ivie 31.

Gelbach, J. B. (2016). When do covariates matter? and which ones, and how much? *Journal of Labor Economics*, 34(2):509–543.

Gibson, J. and McKenzie, D. (2011a). The microeconomic determinants of emigration and return migration of the best and brightest: Evidence from the Pacific. *Journal of Development Economics*, 95(1):18–29.

Gibson, J. and McKenzie, D. (2011b). The microeconomic determinants of emigration and return migration of the best and brightest: Evidence from the Pacific. *Journal of Development Economics*, 95(1):18–29.

Giuliani, D., Dickson, M. M., Espa, G., and Santi, F. (2020). Modelling and predicting the spatio-temporal spread of covid-19 in italy. *BMC infectious diseases*, 20(1):1–10.

Giustinelli, P. (2016). Group decision making with uncertain outcomes: Unpacking child–parent choice of the high school track. *International Economic Review*, 57(2):573–602.

Gong, Y., Stinebrickner, T., Stinebrickner, R., and Yao, Y. (2022). The Role of Non-Pecuniary Considerations: Location Decisions of College Graduates from Low Income Backgrounds. *Centre for Human Capital and Productivity (CHCP) Working Paper Series*, Working Paper #2022-01:58.

González-Leonardo, M., López-Gay, A., and Esteve, A. (2022). Interregional migration of human capital in spain. *Regional Studies, Regional Science*, 9(1):324–342.

Grewenig, E., Lergetporer, P., Werner, K., Woessmann, L., and Zierow, L. (2021). Covid-19 and educational inequality: how school closures affect low-and high-achieving students. *European economic review*, 140:103920.

Gröger, A. (2021). Easy come, easy go? Economic shocks, labor migration and the family left behind. *Journal of International Economics*, 128:103409.

Grogger, J. and Hanson, G. H. (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics*, 95(1):42–57.

Huttunen, K., Møen, J., and Salvanes, K. G. (2018). Job Loss and Regional Mobility. *Journal of Labor Economics*, 36(2):31.

Jaeger, D. A., Arellano-Bover, J., Karbownik, K., Martínez-Matute, M., Nunley, J. M., Seals, A., Alston, M., Becker, S. O., Beneito, P., Böheim, R., et al. (2021). The global covid-19 student survey: First wave results.

Jensen, R. (2010). The (Perceived) Returns to Education and the Demand for Schooling*. *The Quarterly Journal of Economics*, 125(2):515–548.

Joyce, T., Crockett, S., Jaeger, D. A., Altindag, O., and O'Connell, S. D. (2015). Does classroom time matter? *Economics of Education Review*, 46:64–77.

Kaestner, R. and Malamud, O. (2014). Self-Selection and International Migration: New Evidence from Mexico. *The Review of Economics and Statistics*, 96(1):78–91.

Kaufmann, K. M. (2014). Understanding the income gradient in college attendance in Mexico: The role of heterogeneity in expected returns. *Quantitative Economics*, 5(3):583–630.

Kennan, J. and Walker, J. R. (2011). The Effect of Expected Income on Individual Migration Decisions. *Econometrica*, 79(1):211–251.

Koşar, G., Ransom, T., and van der Klaauw, W. (2021). Understanding migration aversion using elicited counterfactual choice probabilities. *Journal of Econometrics*, page S0304407621000415.

Lu, Y. and Song, H. (2020). The effect of educational technology on college students' labor market performance. *Journal of Population Economics*, 33(3):1101–1126.

Maldonado, J. E. and De Witte, K. (2021). The effect of school closures on standardised student test outcomes. *British Educational Research Journal*.

Manski, C. F. (2004). Measuring Expectations. *Econometrica*, 72(5):1329–1376.

McKenzie, D., Gibson, J., and Stillman, S. (2013). A land of milk and honey with streets paved with gold: Do emigrants have over-optimistic expectations about incomes abroad? *Journal of Development Economics*, 102:116–127.

McKenzie, D., Stillman, S., and Gibson, J. (2010). How important is selection? Experimental vs. non-experimental measures of the incomes gains from migration. *Journal of the European Economic Association*, 8(4).

Melguizo, C. and Royuela, V. (2020). What drives migration moves to urban areas in spain? evidence from the great recession. *Regional Studies*, 54(12):1680–1693.

Moretti, E. (2013). *The new geography of jobs*. Houghton Mifflin Harcourt.

OECD (2020). Job creation and local economic development 2020: Rebuilding better. *OECD Publishing, Paris.*

Papke, L. E. and Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of applied econometrics*, 11(6):619–632.

Patt, A., Ruhose, J., Wiederhold, S., and Flores, M. (2021). International emigrant selection on occupational skills. *Journal of the European Economic Association*, 19(2):1249–1298.

Ransom, T. (2022). Ransom, T. (2022). Labor market frictions and moving costs of the employed and unemployed. *Journal of Human Resources*, 57.

Rasch, G. (1993). *Probabilistic models for some intelligence and attainment tests.* ERIC.

Rodríguez-Pose, A. (2018). The revenge of the places that don't matter (and what to do about it). *Cambridge journal of regions, economy and society*, 11(1):189–209.

Romero-Valiente, J. M. and Hidalgo-Capitán, A. L. (2014). El subregistro consular: magnitudes y efectos en las estadísticas de emigración española.

Rosso, A. (2019). Emigrant selection and wages: The case of Poland. *Labour Economics*, 60:148–175.

Sjaastad, L. A. (1962). The costs and returns of human migration. *Journal of political Economy*, 70(5, Part 2):80–93.

Stinebrickner, R. and Stinebrickner, T. R. (2014). A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout. *The Review of Economic Studies*, 81(1):426–472.

Thom, K. (2010). Repeated circular migration: Theory and evidence from undocumented migrants. *New York University, mimeo*.

Tunali, I. (2000). Rationality of Migration. *International Economic Review*, 41(4):893–920.

Vigdor, J. L., Ladd, H. F., and Martinez, E. (2014). Scaling the Digital Divide: Home Computer Technology and Student Achievement. *Economic Inquiry*, 52(3):1103–1119. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecin.12089.

Wiswall, M. and Zafar, B. (2015). Determinants of College Major Choice: Identification using an Information Experiment. *The Review of Economic Studies*, 82(2):791–824.

Wiswall, M. and Zafar, B. (2021). Human Capital Investments and Expectations about Career and Family. *journal of political economy*, 129(5):64.

Zafar, B. (2013). College Major Choice and the Gender Gap. *Journal of Human Resources*, 48(3):52.

Zerecero, M. (2022). The Birthplace Premium. *Working paper*, page 93.

A

Appendix to Chapter 1

A.0.1 Survey administration

Current data protection policies prevent Spanish universities from sharing students' personal information with third parties. Thus, survey administration was carried out with the collaboration of the 5 different departments, who kindly offered to send the survey link through their email account-this is the official channel that is used to communicate news to the students at faculty level-. This procedure to administer the survey necessarily required the survey link included in the email to be an *anonymous link*[1] type. The downside of using this type of survey link is that the survey can be completed multiple times by the same individual. Thus, these links are generally incompatible with monetarily incentivized surveys, a widespread procedure in economics to ensure survey participation -as well as important in this study-.

To keep the environment controlled while ensuring a high participation, I divided the study into two surveys, that were distributed 2 or 3 days apart. Importantly, these surveys were

[1]An anonymous link is an URL link that can be pasted into an email, website or any other mode of communication. In general, a participant can take the survey as many times as she wants. The "Prevent Ballot Box Stuffing" option for anonymous links in Qualtrics (which was used in this survey), deters respondents from taking a survey multiple times by placing a cookie on their browser when they submit a response. However, it can be circumvented by savvy participants clearing their browser cookies, switching to a different web browser, or using a different device.

not presented as independent ones, nor did the second survey come as a surprise. From the beginning, the email that students received from the universities introduced the study as a two survey one. The email described the timing, duration and financial compensation of each survey, and explained that participation in the first survey was necessary but not sufficient to receive the second survey link. To ensure a truthful reporting, however, the selection criterion for the reception of the follow-up survey was not mentioned to the students. The email also explained that all the survey payments would be done through a mobile phone payment platform that is widely used in Spain: BizumAfter the description, the email included the link to the first survey.

The purpose that each survey ought to serve was different, and thus, so were their designs. The goal of the first survey was twofold: First, to collect students' personal information. This allowed not only to send the second survey through an *individual link*[2], but will also allow to contact students in the future. Second, to limit participation in the second survey to those students who were born in Andalusia.

A.0.2 Individual specific migration alternatives

To make the migration alternatives easier to envision, the survey described students' migration alternatives with their reported study plans and migration destination. These are just individual-specific versions of the counterfactual alternatives. For example, if student i answered that her highest expected level of education was a master's degree in *[1. Another Spanish region/2. Province of Madrid]* and that she would move to *[1. Another Spanish region/2. Province of Madrid]* if she were to move outside her region of birth, her migration alternatives were described as follows:

[2]An individual link is a personalized link that can only be used once. The respondent's name, email, and other information from the contact list is automatically saved with their survey data. This means one can track responses in progress and send out reminders.

No-migration, $m = 1$*: Within the 3 years after finishing the bachelor's degree, you pursue your Master's degree in [Madrid], you finish your studies, and start living in Andalusia. 10 years after graduation, you continue living in Andalusia.*

Short-term migration, $m = 2$*: Within the 3 years after finishing the bachelor's degree, you pursue your Master's degree in [Madrid], you finish your studies, and continue living in [Madrid]. You return to Andalusia to live within 10 years after graduation.*

Long-term migration, $m = 3$*: Within the 3 years after finishing the bachelor's degree, you pursue your Master's degree in [Madrid], you finish your studies and continue living in [Madrid]. 10 years after graduation, you continue living in [Madrid].*

Students were also shown a simplified version of these alternatives in a table that stated the locations where they were assumed to be living 3 and 10 years after finishing the bachelor's degree in each migration alternative.

A.0.3 Chosen migration destinations

Map 2a reports the fraction of students that choose each place as a destination if they were to migrate. There are two clear most likely destinations: 52% of students would move to the capital city of Spain, Madrid. 32% would migrate to another country (87% of these to another country in Europe). For comparison, Map 2b shows the locations where those who move out of Andalusia move to, when they are between 23-35 years old, using administrative data (Residential Variation Statistics, yearly averages between 2015-2019). These dataset measures all residential variations with origin and/or destination in Spain within a calendar year. This dataset has two caveats to study migration of highly-educated individuals, but to the best of my knowledge is the best data to analyze internal and international migration jointly out of a place. First, the data does not include information on the individual's educational level. The fact that a high fraction of individuals choose the islands as migration destinations (see Map 2b) is driven by those with no tertiary education "Balearic Islands attract large groups of people, but mainly low-educated individuals to the tourist industries" (González-Leonardo et al. (2022), page 237). Second, given that only individuals who register themselves in the new residential location are included in this statistic, the dataset is known to underestimates international migration by about 17%-35% Romero-Valiente and Hidalgo-Capitán (2014). Taking these two issues into account, the reported destinations by survey participants and observed moves in the administrative data follow the same pattern: 27% of young Andalusians move to the capital city of Spain, Madrid and 16% move to a different country.

A.0.4 Migration alternative approximations

For each individual and each migration alternative, the survey elicited beliefs about the expected minimum and maximum full-time earnings, the probability of working full-time, working part-time, having a good study-job match and enjoying social life 3 and 10 years after finishing the bachelor's degree. Using these two points per individual i and migration alternative m, I assume that outcomes have a linear growth throughout the migration alternative. This resembles the growth in career related outcomes typically observed over the very beginning of individuals' labor market trajectory, which is the period in which our migration alternatives are defined. On average, students are 25 years old at the beginning of the alternative, t_{i0}, and are 33 at the end of the alternative, at $T = 10$. Specifically, we approximate each alternative as follows.

By definition of the alternatives, the no-migration and long-term migration alternatives imply living in the same location (in their region of birth, Andalusia, and at each students' chosen migration destination respectively) throughout the entire alternative. For these two alternatives, we assume that these outcomes grow linearly throughout the alternative and denote the growth rates as g_{i,x,m_1} and g_{i,x,m_3}, where the subscripts i and x denote that growth rates are individual and outcome specific, and m_1 and m_3 refer to the no-migration and long-term migration alternatives respectively.

$$
\begin{aligned}
g_{i,x,m_1} &= \frac{Pr(x|i,m=1,t=3) - Pr(x|i,m=1,t=10)}{10-3} \\
g_{i,x,m_3} &= \frac{Pr(x|i,m=3,t=3) - Pr(x|i,m=3,t=10)}{10-3}
\end{aligned} \tag{A.1}
$$

where x is equal to minimum and maximum earnings if working full-time, study-job match, working full-time, working part-time and enjoying social life. Then, the probability of outcome x in period t in migration alternatives m_1 and m_3 for student i are defined as:

$$
\begin{aligned}
Pr(x|i,m=1,t) &= Pr(x|i,m=1,t=3) + g_{i,x,m_1} * (t-3), \quad \text{for} \quad t \in [t_{i0}, 10] \\
Pr(x|i,m=3,t) &= Pr(x|i,m=3,t=3) + g_{i,x,m_3} * (t-3), \quad \text{for} \quad t \in [t_{i0}, 10]
\end{aligned}
\tag{A.2}
$$

where the value of t_{i0} depends on the expected maximum level of education of student i, with t_{i0} being equal to 0,1, 2 and 3 for students whose maximum expected level of education is a bachelor's degree, other type of studies, a master's degree and a PhD respectively.

The survey did not ask students about the period in which they would return back to their region of birth to work conditional on choosing the short-term migration trajectory. I assume that all students return back at period $t = 7$, which corresponds to living at the chosen destination for 4 years. This is roughly the average time that internal southern migrants in Spain spent in the north, as shown in Table A.2. Thus, in the short-term migration alternative, $m = 2$, students are assumed to be living at their chosen destination $\forall t, t < 7$, and back at their region of birth at $\forall t, t \geq 7$. For this alternative, we assume that students believe each outcome's growth rate to be location specific, and approximate the alternative using the previously calculated individual and location-specific growth rates, g_{i,x,m_1} and g_{i,x,m_3}. The probability of outcome x in period t for student i in the short-term migration alternative, $m = 2$, is defined as:

$$
Pr(x|i,m=2,t) = \begin{cases} Pr(x|i,m=2,t=3) + g_{i,x,m_3} * (t-3), & \text{for} \quad t_{i0} \leq t \leq 6 \\ Pr(x|i,m=2,t=10) + g_{i,x,m_1} * (t-10), & \text{for} \quad 6 < t \leq 10 \end{cases}
\tag{A.3}
$$

where the value of t_{i0} is defined as above.

Finally, students were asked about the probability of enjoying being close to family members, partner and friends 3 years after finishing the bachelor's degree only, if they were living in their region of birth and if they were living at their chosen migration destination. We assume that the value of this outcome is location specific and constant over time. Then, probability of enjoying being close to family members, partner and friends in period t in migration alternatives m_1 and m_3 for student i are defined as:

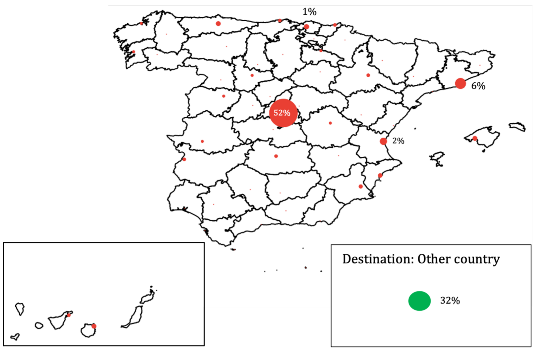
$$\begin{aligned} Pr(x|i,m=1,t) &= Pr(x|i,m=1,t=3), \quad \text{for} \quad t \in [t_{i0},10] \\ Pr(x|i,m=3,t) &= Pr(x|i,m=3,t=3), \quad \text{for} \quad t \in [t_{i0},10] \end{aligned} \tag{A.4}$$

where the value of t_{i0} is defined as above. And the probability in migration alternative m_2 as:

$$Pr(x|i,m=2,t) = \begin{cases} Pr(x|i,m=3,t=3), & \text{for} \quad t_{i0} \leq t \leq 6 \\ Pr(x|i,m=1,t=3), & \text{for} \quad 6 < t \leq 10 \end{cases} \tag{A.5}$$

Figure A.1 Chosen destinations

(a) Survey



(b) Residential Variation Statistics

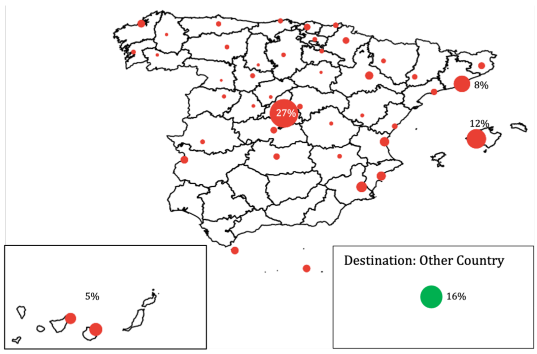
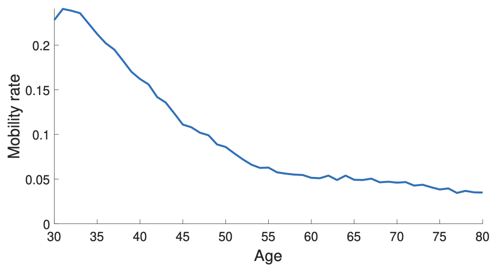


Fig 2.a.: Chosen destinations by survey participants: 52% Madrid, 28% Europe, 6% Barcelona, 2% North America, 1% South America, 1% Biscay. Remaining destinations, less than 1% each. Fig 2.b.: Migration moves using administrative data (Residential Variation Statistics). Sample restricted to individuals born in Andalusia, that move out of this region when they are between 23-35 years old. 27% Madrid, 16% another country, 12% Balearic Islands, 8% Barcelona, 5% Canary Islands.

Figures

Figure A.1 Migration hazard by age



(c) Mean mobility rate by age

Notes: The top left panel displays the average urban area unemployment rate across all individuals arriving (separating from) an urban area. The top right panel displays the arrival rate of people in an urban area depending on its tercile in the urban area unemployment distribution relative to the arrival rate in the lowest tercile. The bottom panel shows the mean decennial mobility rate of individuals over age. Source: 1991, 2001, and 2011 Censuses.

Notes:

Figure A.2 Location of universities in the sample

Note: The university marked in red is the University of Seville (Universidad de Sevilla, U.S.) and the one marked in blue is the University of Granada (Universidad de Granada, UGR). The UGR is located in Granada -the capital city of the province of Granada- and the US is located in Seville -both the capital city of the province of Seville and of Andalusia-. The blue line marks the borders of the region of Andalusia.

Figure A.3 Kernel densities of migration choice probabilities across students

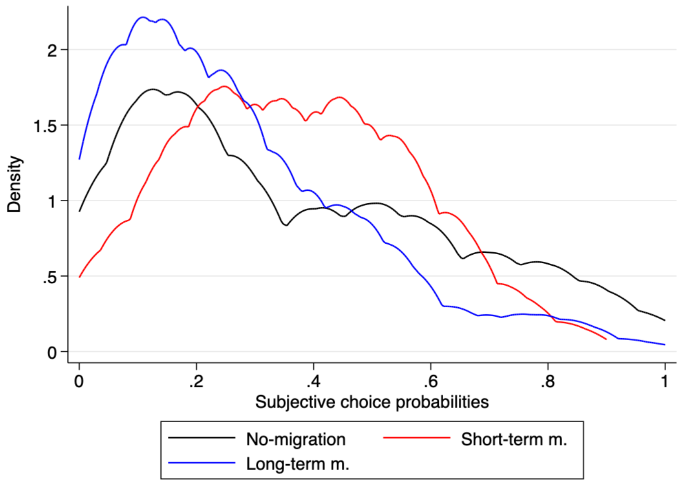
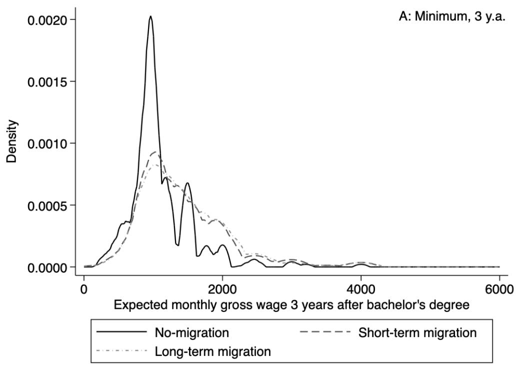
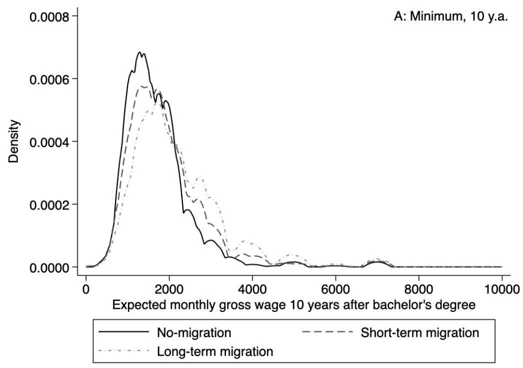


Figure A.4 Probability density function of expected minimum monthly gross earnings 3 and 10 years after finishing the bachelor's degree by migration alternative

(a) 3 years after



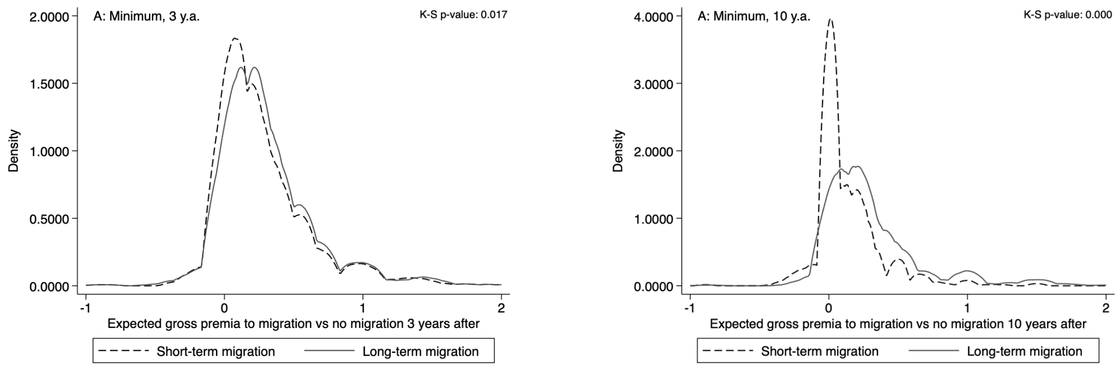
(b) 10 years after



Note:

Figure A.5 Minimum earnings premia from short-term and long-term migration relative to no-migration

(a) 3 years after **(b)** 10 years after



Note: Premia from short-term vs no-migration and from long-term vs no-migration. K-S = Kolmogorov–Smirnov.

Figure A.6 Distribution of full-time earnings premium from short-term migration relative to no - migration 10 years after

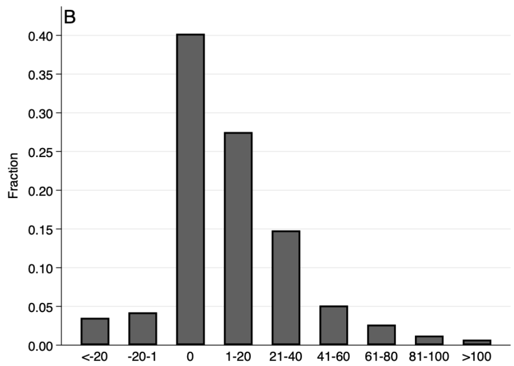
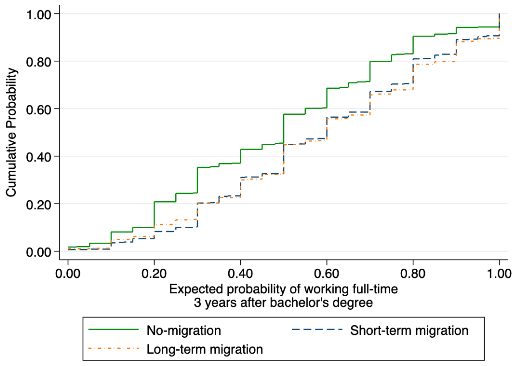
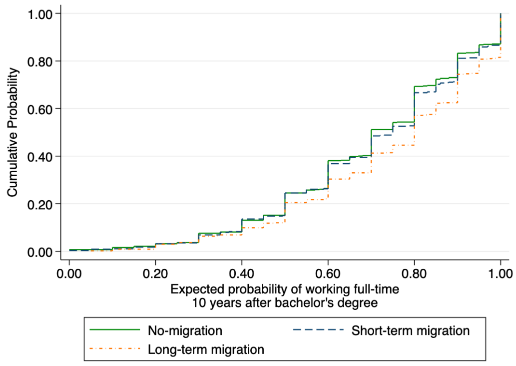


Figure A.7 Probability of working full-time 3 and 10 years after by migration alternative

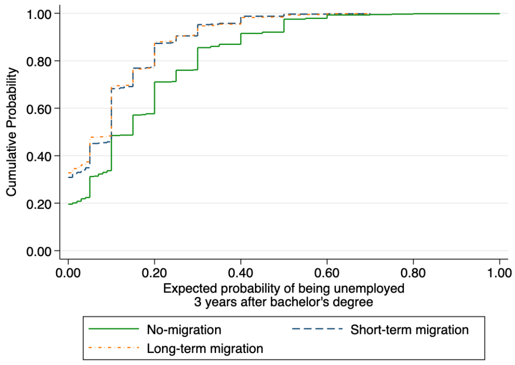
(a)



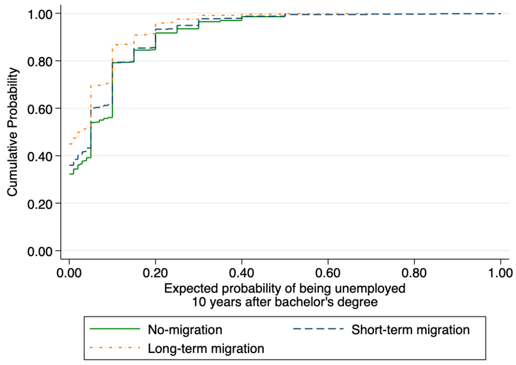
(b)



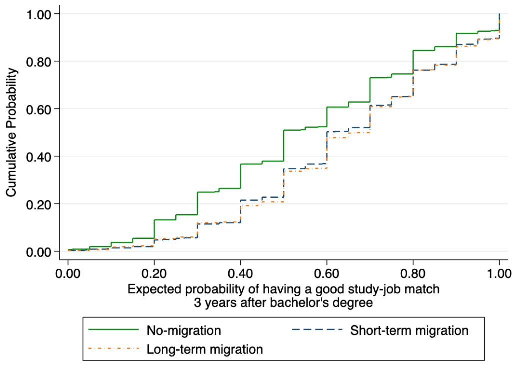
(c)



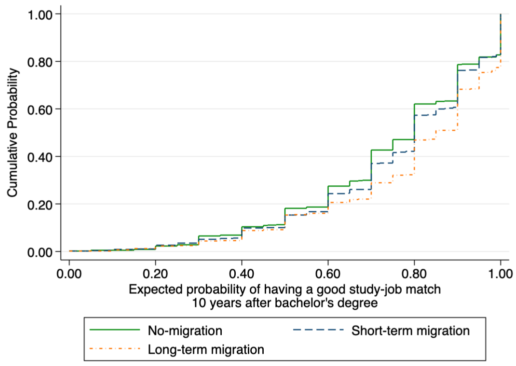
(d)



(e)



(f)



Note:

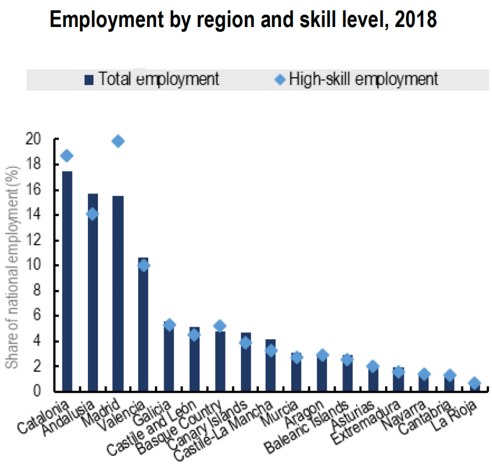
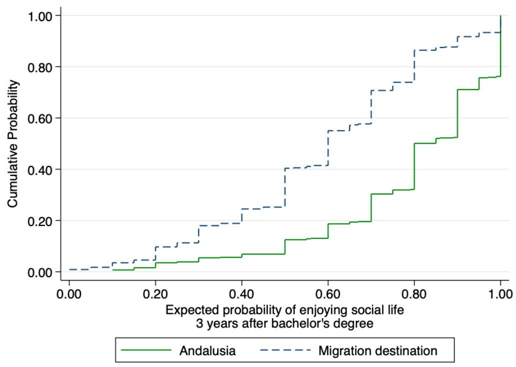


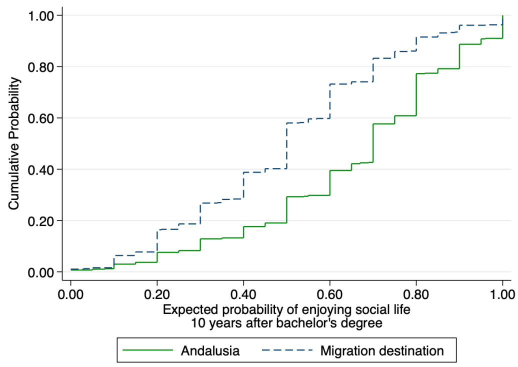
Figure A.8 Source: OECD (2020), Job Creation and Local Economic Development 2020. Rebuilding Better, OECD.

Figure A.9 Probability of enjoying social life 3 and 10 years after by location

(a)

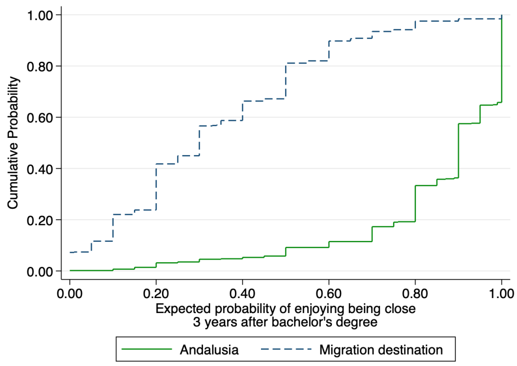


(b)



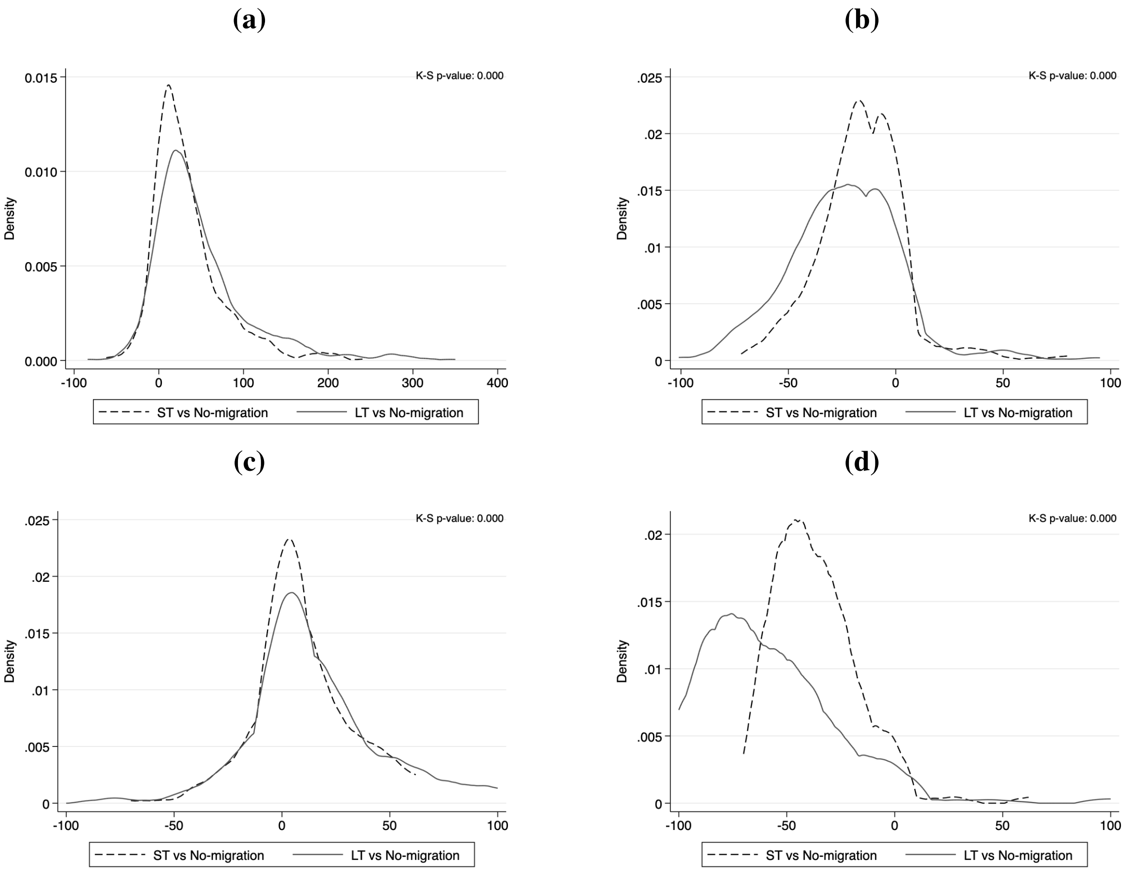
}

(c)



Note:

Figure A.10 Densities of life-cycle premia



Note:

Tables

Table A.1 Sorting into migration paths of cohorts born in any region in the south and the southern region of Andalusia between 1980-1984, by level of education.

	No tertiary education		Tertiary education	
	Andalusia	Any, south	Andalusia	Any, south
	(1a)	(1b)	(2a)	(2b)
Stayers (%)	84.29	83.78	71.32	70.74
Short-term migrants (%)	13.6	13.86	20.05	21.2
Long-term migrants (%)	2.11	2.36	8.64	8.05
N	11,969	29,963	3,231	10,206

Note: Data from the Spanish Work History Sample, waves 2006-2019. The sample is restricted to individuals born born between 1980-1984. Migration moves are identified by the location of the establishment where the individual works. Only first out-moves and first return moves are identified. Individuals are followed from their first employment until the year 2019. Columns 1a and 2a refer to individuals who are born and start working in Andalusia. Columns (2a) and (2b) are individuals who are born and start working in the south of Spain.

Table A.2 Length of stay in the north for cohorts born 1980-1985, with tertiary education and that started working in the south

	Long-term migrants		Short-term migrants	
	Andalusia	Any, south	Andalusia	Any, south
Age at labor market entry	22.07	21.79	21.25	21.23
Age at first move to the north	29.56	29.16	25.87	26.01
Number of months in the north	76.78	79.19	48.91	44.49

Note: Data from the Spanish Work History Sample, waves 2006-2019. The sample is restricted to individuals born in the south of Spain and born between 1980-1984. Migration moves are identified by the location of the establishment where the individual works. Only first out-moves and first return moves are identified. Individuals are followed from their first employment until the year 2019. All individuals have tertiary education.

Table A.3 Potential sources of information

	Mean
Own labor market experience while studying	
Has no experience	0.58
Job search	
Has not applied for a job	0.68
Has applied for a job	
- In Andalusia	0.78
- In another Spanish region	0.14
- In another country	0.08
Siblings' labor market experience	
Has no older sibling working	0.62
Has older sibling working	
- In Andalusia	0.67
- In another Spanish region	0.18
- In another country	0.15

Table A.4 Categorization of Degrees

Degrees in the Department of Business and Economics
Double Degree in Business Administration and Management and Law
Double Degree in Business Administration and Management and Building
Double Degree in Law and Economics
Degree in Business Administration and Management
Degree in Economics
Degreen in Tourism
Degree in Marketing and Market Research
Degree in Finance and Accounting
Degrees in the Department of Engineering
Degree in Aerospace Engineering
Degree in Civil Engineering
Degree in Electronic, Robotics and Mechatronics Engineering
Degree in Industrial Organization Engineering
Degree in Energy Engineering
Degree in Industrial Technologies Engineering
Degree in Telecommunication Technologies Engineering
Degrees in the Department of Natural Sciences
Double Degree in Computer Engineering and Mathematics
Degree in Industrial Electronic Engineering
Degree in Chemical Engineering
Degree in Biology
Degree in Biochemistry
Degree in Biotechnology
Degree in Statistics
Degree in Physics
Degree in Geology
Degree in Mathematics
Degree in Chemistry
Degree in Optics and Optometry

Table A.5 Short-term and log-term migration premia relative to no-migration

	3 years after		10 years after	
	ST vs Stay.	LT vs Stay	ST vs Stay	LT vs Stay
	Panel A. Minimum full-time earnings			
Average	0.29***	0.32***	0.13***	0.31***+++
Median	0.20	0.25	0.06	0.20
Standard Deviation	0.41	0.42	0.23	0.40
	Panel B. Full-time employment			
Average	0.47***	0.49***	0.06***	0.13***+++
Median	0.14	0.14	0.00	0.05
Standard Deviation	1.21	1.33	0.44	0.50
	Panel C. Part-time employment			
Average	-0.03	-0.04*	0.14***	0.09
Median	0.00	0.00	-0.00	-0.00
Standard Deviation	0.60	0.59	0.91	1.14
	Panel D. Unemployed			
Average	-0.28***	-0.25***	0.01	-0.18*
Median	-0.40	-0.50	0.00	-0.50
Standard Deviation	0.68	1.04	1.25	1.97
	Panel E. Minimum earnings (weighted earnings)			
Average	0.65***	0.69***	0.19***+++	0.44 ***+++
Median	0.33	0.37	0.09	0.28
Standard Deviation	1.28	1.27	0.72	0.67
	Panel F. Good study-job match			
Average	0.41***	0.46***	0.09***	0.18***
Median	0.14	0.15	0.00	0.00
Standard Deviation	1.13	1.42	0.74	1.12

*** Means statistically significant at the 1% level
+++ Means of ST and LT premia statistically significant at the 1% level.

Table A.6 Location premia 3 and 10 years after

	Social Life		Enjoy Close
	3 years after	10 years after	3 years after
Average	-0.20***	-0.18***	-0.56***
Median	-0.23	-0.20	-0.62
Standard Deviation	0.47	0.48	0.44

*** Means statistically significant at the 1% level

Table A.7 Ex-ante premia from short-term migration and long-term migration relative to no-migration

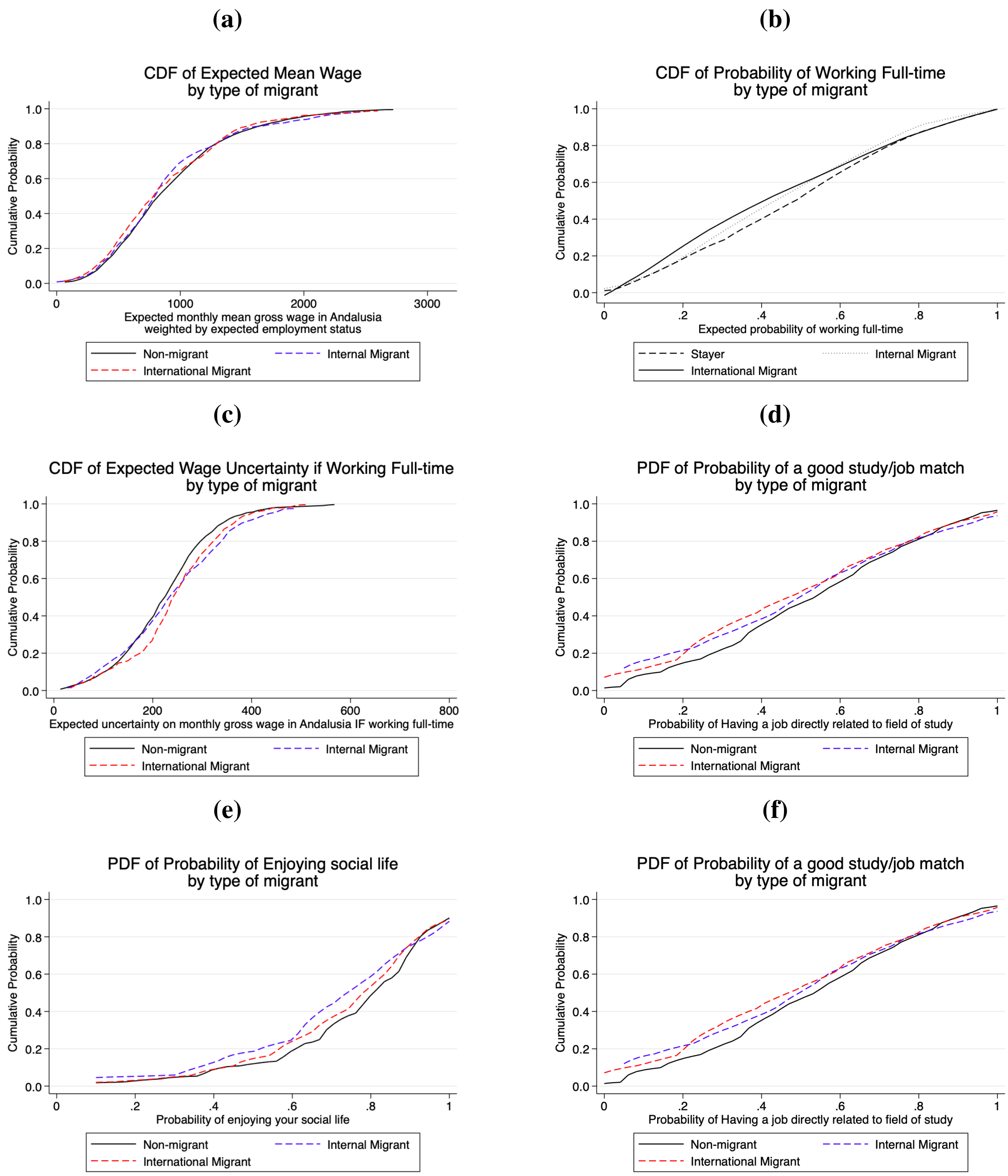
	ST vs No-migration	LT vs No-migration
Unconditional earnings	37.26	53.02
	[-2.89; 23.39; 93.82]	[-2.03; 33.92; 127.06]
	(54.42)	(73.60)
Study-job match	20.91	29.69
	[-15.13; 7.75; 57.63]	[-15.83; 10.81; 77.28]
	(28.17)	(34.39)
Enjoy Social Life	-14.76	-22.24
	[-39.98; -15.55; 0]	[-56.63; -23.42; 0]
	(23.91)	(32.76)
Enjoy Being close	-35.02	-55.77
	[-61.18; -39.77; -8.74]	[-55.78; -62.5; -14.28]
	(27.68)	(44.10)

The first row of each outcome reports the mean of the distribution. 10th, 50th and 90th percentiles are reported in brackets and standard deviations in parenthesis.

B

Appendix to Chapter 2

Figure B.1



Note:

C

Appendix to Chapter 3

Table C.1 Online Learning Mandates 2020-2021 School Year

Dates	Zone	*Scuola Primaria* Grades 1-5	*Scuola Secundaria di Primo Grado* Grade 6	Grades 7 and 8	*Scuola Secundaria di Secondo Grado* Grades 9-13
November 6, 2020 - January 6, 2021 (DPCM November 3, 2020)	Yellow	in-person	in-person	in-person	e-learning
	Orange	in-person	in-person	in-person	e-learning
	Red	in-person	in-person	e-learning	e-learning
January 7 - March 5, 2021 (DL January 5, 2021)	Yellow	in-person	in-person	in-person	50-75% in-person
	Orange	in-person	in-person	in-person	50-75% in-person
	Red	in-person	in-person	e-learning	e-learning
March 5 - School end (DPCM March 2, 2021)	Yellow	in-person	in-person	in-person	50-75% in-person
	Orange	in-person	in-person	in-person	50-75% in-person
	Red	e-learning	e-learning	e-learning	e-learning

Note: This table reports the changes in the online learning mandates that took place during the 2020-2021 school year. “50-75% in-person" means that 50 to 75 percent of the students were allowed to attend in-person lessons. From March 5, 2021 regions and autonomous provinces were allowed to impose a color increase within their territories if specific epidemiological conditions were met. Implying that different colors could be imposed within a region.

Table C.2 Results of Before-After Analysis on Google Search Index: Alternative School Closures

	(1) All	(2) All	(3) GC	(4) GC	(5) WS	(6) WS	(7) Ed	(8) Ed	(9) Sc	(10) Sc	(11) St	(12) St
After 4 March	1.242***		2.714***		2.537***		1.974***		-0.634*		-0.551*	
	(0.201)		(0.321)		(0.314)		(0.302)		(0.380)		(0.288)	
Before 15 Feb. after 15 Mar.		1.956***		4.131***		3.825***		2.963***		-0.901*		-0.580
		(0.259)		(0.132)		(0.123)		(0.162)		(0.496)		(0.379)
North	0.076***	0.088***	0.215***	0.229***	0.191***	0.204***	0.216***	0.228***	-0.155***	-0.148**	-0.091**	-0.076*
	(0.018)	(0.018)	(0.013)	(0.013)	(0.014)	(0.014)	(0.017)	(0.016)	(0.058)	(0.059)	(0.045)	(0.045)
ln(COVID-19 Cases)	0.085***	0.005	0.056	-0.100***	0.099***	-0.044***	0.032	-0.079***	0.126***	0.150***	0.128***	0.135***
	(0.023)	(0.030)	(0.037)	(0.016)	(0.036)	(0.014)	(0.035)	(0.019)	(0.043)	(0.055)	(0.033)	(0.042)
Share of Internet Access	0.015***	0.015***	0.020***	0.021***	-0.002	-0.001	0.010***	0.011***	0.023***	0.022***	0.025***	0.024***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.006)	(0.006)	(0.005)	(0.005)
Constant	0.172	0.125	-0.835***	-0.952***	0.619***	0.520***	0.769***	0.681***	0.083	0.109	0.243	0.293
	(0.150)	(0.150)	(0.115)	(0.105)	(0.117)	(0.110)	(0.145)	(0.138)	(0.473)	(0.479)	(0.375)	(0.378)
Observations	19,776	19,392	4,120	4,040	4,120	4,040	4,120	4,040	3,708	3,636	3,708	3,636
Adjusted R-squared	0.481	0.485	0.884	0.893	0.877	0.884	0.804	0.810	0.218	0.219	0.239	0.243
Platform FEs	Yes	Yes	-	-	-	-	-	-	-	-	-	-
Academic year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week of the year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

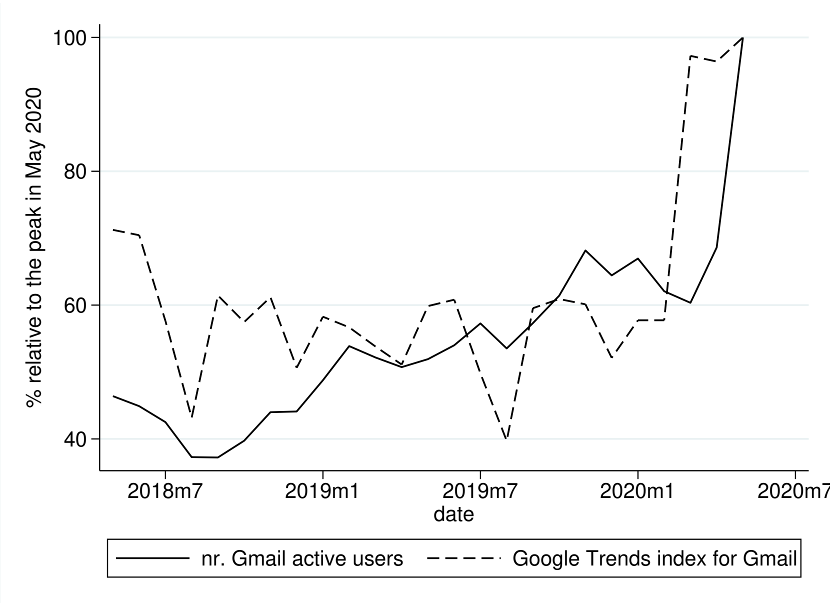
Notes: This table reports the results from estimating equation **??** by ordinary least squares during the period of June 27th 2016 to June 7th 2020. The dependent variable is the logarithm of the Google Search Index for selected E-learning platforms. *After March 4* takes value 1 after March 4 2020 and 0 before. *Before 15 Feb. After 15 March* takes value 1 after March 15 2020 and 0 before 15 February. *North* takes value 1 for Emilia-Romagna and all regions above it, and 0 otherwise. *Share of Internet Usage* contains the share of households in each region that had internet access in 2019. *ln(COVID-19 Cases)* contains the total number of COVID-19 cases reported in each region and day. GC stands for Google Classroom, WS for WeSchool, Ed for Edmodo, Sc for Scuola.net and St for Studenti.it. All regression coefficients are weighted by each region's population and include fixed effects for each searched platform, academic year and week of year. Heteroskedasticity robust standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table C.3 Difference-in-Difference Results: Alternative School Closures

VARIABLES	(1) All	(2) All	(3) GC	(4) GC	(5) WS	(6) WS	(7) Ed	(8) Ed	(9) Sc	(10) Sc	(11) St	(12) St
INVALSI Score	0.067	0.066	0.099	0.093	0.062	0.057	0.051	0.047	0.095	0.107	0.029	0.024
	(0.201)	(0.203)	(0.095)	(0.097)	(0.047)	(0.048)	(0.102)	(0.103)	(0.536)	(0.532)	(0.484)	(0.485)
After 4 March	0.752***		1.959***		1.984***		1.625***		-0.851		-1.176***	
	(0.197)		(0.229)		(0.259)		(0.304)		(0.539)		(0.454)	
After 4 March * INVALSI Score	-0.216***		-0.335***		-0.244***		-0.156**		-0.103		-0.266***	
	(0.044)		(0.054)		(0.063)		(0.062)		(0.076)		(0.089)	
Before 15 Feb. after 15 Mar.		1.435***		3.564***		3.705***		3.122***		-1.939**		-2.045**
		(0.457)		(0.204)		(0.239)		(0.415)		(0.969)		(0.969)
Before 15 Feb. after 15 Mar. * INVALSI Score		-0.126**		-0.141***		-0.034		0.032		-0.244*		-0.330**
		(0.056)		(0.035)		(0.034)		(0.072)		(0.133)		(0.162)
North	0.002	0.010	0.107	0.116	0.125	0.134*	0.159	0.169	-0.269	-0.276	-0.115	-0.099
	(0.248)	(0.250)	(0.131)	(0.133)	(0.078)	(0.080)	(0.193)	(0.196)	(0.651)	(0.645)	(0.622)	(0.621)
ln(COVID-19 Cases)	0.146***	0.066	0.150***	-0.034	0.167***	-0.030	0.076**	-0.098*	0.153**	0.270**	0.206***	0.305***
	(0.022)	(0.052)	(0.024)	(0.025)	(0.026)	(0.026)	(0.037)	(0.050)	(0.067)	(0.111)	(0.053)	(0.109)
Share of Internet Access	0.009	0.008	0.010	0.011	-0.008	-0.007	0.005	0.005	0.012	0.011	0.023	0.023
	(0.027)	(0.027)	(0.014)	(0.014)	(0.007)	(0.008)	(0.019)	(0.019)	(0.075)	(0.074)	(0.068)	(0.068)
Constant	0.691	0.690	-0.081	-0.141	1.074*	1.032*	1.164	1.124	0.923	1.024	0.380	0.395
	(2.081)	(2.108)	(1.075)	(1.098)	(0.567)	(0.589)	(1.428)	(1.448)	(5.743)	(5.686)	(5.238)	(5.228)
Observations	19,776	19,776	4,120	4,120	4,120	4,120	4,120	4,120	3,708	3,708	3,708	3,708
Platform FEs	Yes	Yes	-	-	-	-	-	-	-	-	-	-
Academic year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week of the year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

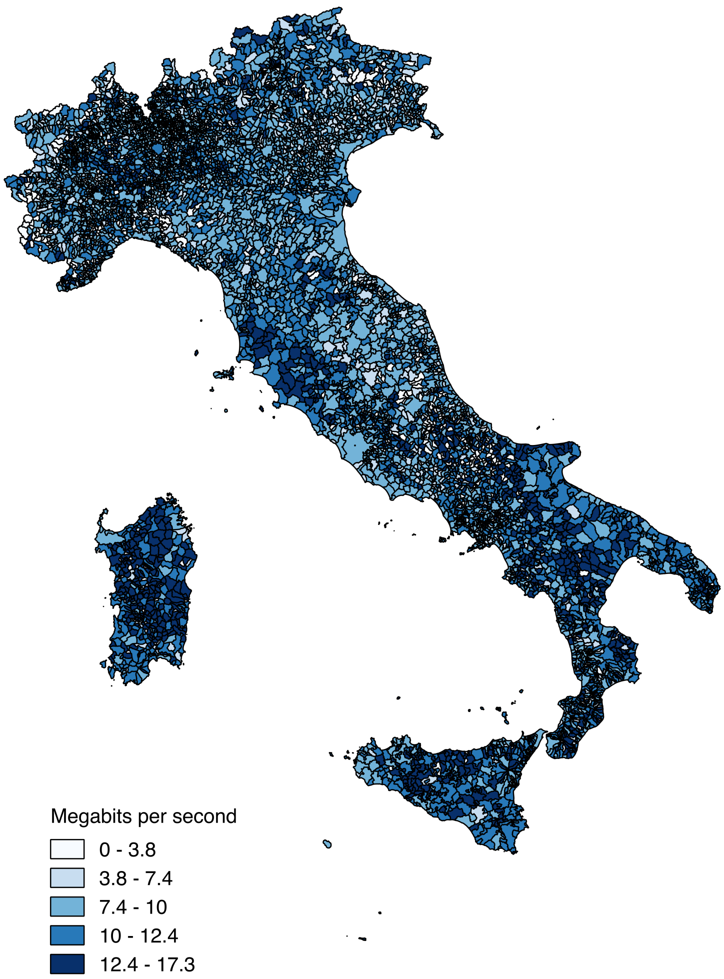
Notes: This table reports the results from estimating equation 3.1 by ordinary least squares during the period June 27th 2016 to June 7th 2020. The dependent variable is the logarithm of the Google Search Index for selected E-learning platforms. *After March 4* takes value 1 after March 4 2020 and 0 before. *Before 15 Feb. After 15 March* takes value 1 after March 15 2020 and 0 before 15 February. *INVALSI Score* represents the average score obtained in 2018 in the INVALSI test for Italian language. This variable has been standardised (demeaned and divided by its standard deviations) hence its units are standard deviations. *North* takes value 1 for Emilia-Romagna and all regions above it, and 0 otherwise. *Share of Internet Usage* contains the share of households in each region that had internet access in 2019. *ln(COVID-19 Cases)* contains the total number of COVID-19 cases reported in each region and day. GC stands for Google Classroom, WS for WeSchool, Ed for Edmodo, Sc for Scuola.net and St for Studenti.it. All regression coefficients are weighted by each region's population and include fixed effects for each searched platform, academic year and week of year. Bootstrapped standard errors are clustered by region and reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Figure C.1 Comparison between number of active Gmail users and Google Trends Index for Gmail



Note: This figure plots the average monthly number of active users of Gmail, provided by AirnowData, and the average monthly Google Trends index for Gmail, between May 2018 and May 2020. Both series are rescaled relative to the peak in May 2020.

Figure C.2 Geographic Distribution of ADSL Download Speed in 2018



Note: This figure plots the average ADSL download speed in each Italian municipality in December 2018. Lighter colours indicate no data or low download speeds while darker colours represent higher average download speeds. Source: Autorità per le Garanzie nelle Comunicazioni (AGCOM).

Table C.4 Descriptive Statistics on the Colour System during Schooling Days of 2020/2021

Region	First Date Red Zone	Last Date Red Zone	Nr of Times Red Zone	Share of Days in Red	Orange	Yellow	White
Abruzzo	22/11/2020	05/04/2021	5	.16	.52	.31	.01
Apulia	24/12/2020	25/04/2021	4	.24	.35	.41	0
Basilicata	24/12/2020	05/04/2021	5	.13	.43	.45	0
Calabria	06/11/2020	11/04/2021	4	.22	.39	.4	0
Campania	15/11/2020	18/04/2021	5	.34	.2	.47	0
Emilia-Romagna	24/12/2020	11/04/2021	4	.18	.39	.43	0
Friuli-Venezia Giulia	24/12/2020	11/04/2021	4	.18	.29	.49	.04
Lazio	24/12/2020	05/04/2021	5	.13	.21	.66	0
Liguria	24/12/2020	05/04/2021	4	.06	.44	.49	.01
Lombardy	06/11/2020	11/04/2021	5	.32	.29	.4	0
Marche	24/12/2020	05/04/2021	4	.15	.35	.5	0
Molise	24/12/2020	05/04/2021	5	.16	.21	.59	.04
Piedmont	06/11/2020	11/04/2021	4	.28	.29	.42	0
Sardinia	24/12/2020	02/05/2021	5	.16	.34	.36	.14
Sicily	24/12/2020	05/04/2021	5	.13	.51	.36	0
Trentino-Alto Adige	10/11/2020	06/04/2021	6	.34	.38	.28	0
Tuscany	15/11/2020	11/04/2021	5	.21	.38	.41	0
Umbria	24/12/2020	05/04/2021	4	.06	.6	.33	.01
Valle d'Aosta	06/11/2020	09/05/2021	5	.35	.33	.33	0
Veneto	24/12/2020	06/04/2021	4	.15	.25	.59	.01

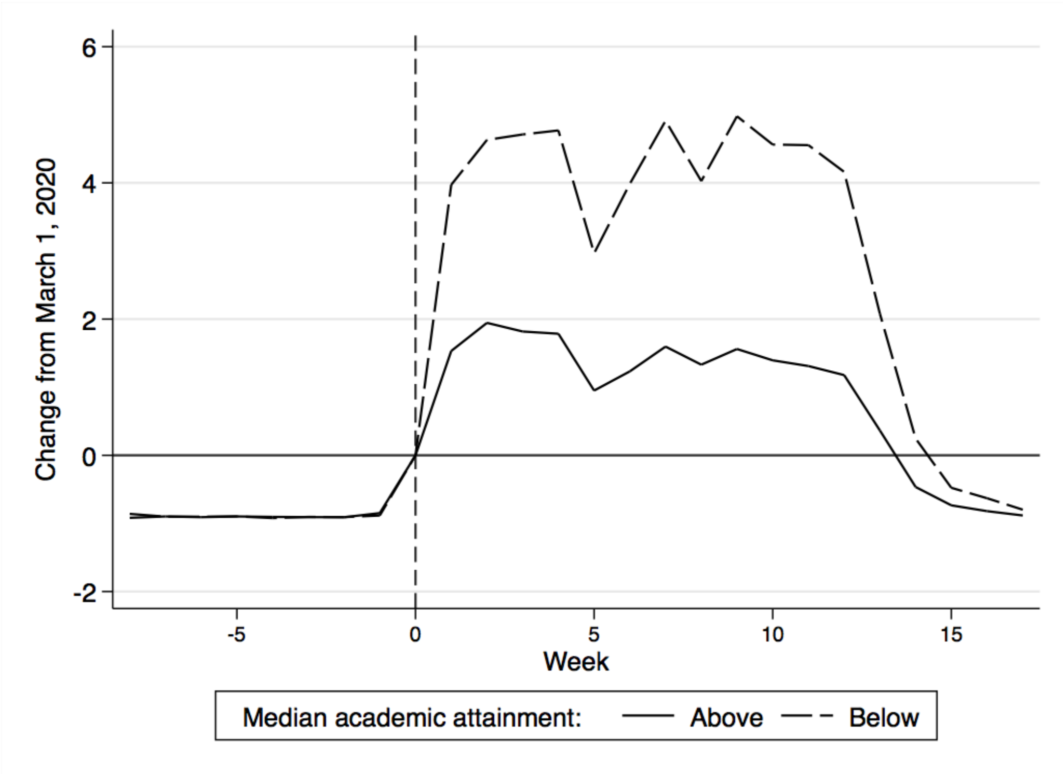
Note: This table reports the descriptive statistics of the colour system in Italian regions between November 6, 2020 and June 8, 2021. Trentino-Alto Adige takes the highest colour in the scale of the two autonomous provinces of Bolzano and Trento in order to make it compatible with the Google Trends data.

Table C.5 Difference-in-Differences Results by Platform - Academic Year 2020/2021

	(1) GC	(2) GC	(3) WS	(4) WS	(5) Ed	(6) Ed
INVALSI Score	0.029	0.036	0.153	0.138	-0.171	-0.188
	(0.110)	(0.104)	(0.143)	(0.155)	(0.349)	(0.349)
Orange	0.093**	0.088**	0.168**	0.169**	0.029	0.021
	(0.037)	(0.036)	(0.071)	(0.076)	(0.086)	(0.094)
Red	0.228***	0.226**	0.522***	0.522***	0.285	0.280
	(0.078)	(0.090)	(0.158)	(0.176)	(0.189)	(0.207)
INVALSI Score x Orange		-0.043		0.021		-0.042
		(0.027)		(0.084)		(0.080)
INVALSI Score x Red		0.017		0.045		0.132
		(0.057)		(0.143)		(0.131)
North	-0.322***	-0.321***	-1.053***	-1.054***	0.155	0.157
	(0.113)	(0.111)	(0.251)	(0.253)	(0.499)	(0.499)
ln(COVID-19 Cases)	0.326***	0.327***	0.810***	0.809***	0.902***	0.901***
	(0.061)	(0.061)	(0.078)	(0.079)	(0.138)	(0.138)
Share of Internet Access	-0.012	-0.011	-0.082***	-0.082***	-0.020	-0.018
	(0.017)	(0.016)	(0.023)	(0.023)	(0.055)	(0.056)
Constant	0.424	0.361	-0.695	-0.695	-7.745*	-7.871*
	(1.375)	(1.351)	(1.775)	(1.800)	(4.098)	(4.123)
Observations	2,720	2,720	2,720	2,720	2,720	2,720
Week of the year FEs	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results from estimating equation (3.2) and (3.3) for each platform. The sample used contains daily observations from November 6th 2020 to June 7 2021, except for weekends and national holidays. The dependent variable is the logarithm of the Google Search Index for *Google Classroom*. *Red Zone* and *Orange Zone* take value 1 when a region is, respectively, red or orange zone in a certain day and 0 otherwise. *INVALSI Score* contains the regional average score of the 2019 INVALSI test in Italian. *North* takes value 1 for Emilia-Romagna and all regions above, and 0 otherwise. *Share of Internet Usage* contains the share of households in each region that used internet in 2019. *COVID-19 Cases* contains the total number of COVID-19 cases reported in each region and day. GC stands for Google Classroom, WS for WeSchool, Ed for Edmodo. Bootstrapped standard errors are clustered by region and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figure C.3 Google Trends Search Index for Google Classroom
by Academic Performance



Note: This figure plots weekly changes of the Google Trends search index for the term *Google Classroom* in two groups of regions relative to March 1, 2020. Search index represented under below (above) the 2019 median INVALSI score contain the population weighted mean of the search index for the regions with a score in Italian below (above) the national median. Regional mean scores in Italian are extracted from the 2019 INVALSI report corresponding to Grade 10 students. Regional population shares used for the weights correspond to 2019 and are extracted from ISTAT.