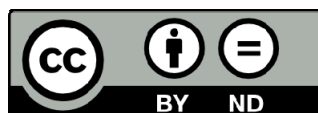




UNIVERSITAT DE
BARCELONA

Essays on Education Economics

Àlex Izquierdo Nofuentes



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- SenseObraDerivada 4.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - SinObraDerivada 4.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-NoDerivatives 4.0. Spain License.**

PhD in Economics

Essays on Education Economics

Àlex Izquierdo Nofuentes



UNIVERSITAT DE
BARCELONA

PhD in Economics

Thesis title:

Essays on Education
Economics

PhD candidate:

Àlex Izquierdo Nofuentes

Advisors:

Andreu Arenas Jal

Daniel Montolio Estivill

Date:

September 2025



UNIVERSITAT DE
BARCELONA

Abstract

This dissertation explores how peer interactions in schools and labor-market institutions jointly shape educational outcomes and the returns to schooling. The first two chapters employ a large-scale lab-in-the-field experiment in schools in the Barcelona Metropolitan Area to examine how students and teachers decide to form groups in class, and how partner characteristics influence both academic performance and subjective well-being. The findings document a systematic trade-off between comfort and efficiency: students choose working with friends and same-gender peers, and working with them increases their reported level of well-being. However, such groupings may lower test performance relative to more heterogeneous teams, and teacher-formed groups are more inclusive, improving the integration of marginalized students. The third chapter turns to labor markets, exploiting staggered judicial reforms in U.S. states that increased dismissal costs for employers. The results show that stronger employment protection raises the wage premium to formal schooling, especially among younger cohorts, but does not generate detectable increases in attained education. Together, the chapters link micro-level peer dynamics in classrooms to macro-level institutional forces in labor markets, highlighting how social preferences and institutional rules reallocate both opportunities and rewards across the life cycle.

Keywords — Education, Group Formation, Peer Effects, Friendship, Performance, Well-being, Labor Market, Dismissal Costs

Acknowledgments

Life is a journey in which we rarely walk alone. Therefore, I believe every personal success must be shared with those who have walked by our side along the way, especially in this case. I cannot imagine myself writing these lines without the accompanying steps of so many people. I am deeply grateful to everyone who has helped me reach this point in my journey.

First and foremost, my advisors, Andreu and Daniel, have been an essential part of this path. I am very grateful for their guidance, patience, and confidence. Their dedication and suggestions made it possible for me to come this far. My gratitude also extends to Pol Campos i Mercadé, with whom I had the pleasure of working during my research stay at Lund University. I greatly enjoyed my time there and learned a lot. Thank you for your many pieces of advice.

Throughout these years, the Institut d’Economia de Barcelona (IEB) and the Departament d’Economia de la Universitat de Barcelona have provided me with a wonderful, friendly, and welcoming environment to learn and grow, but that would not have been possible without the people in it. Within the IEB, I particularly want to thank Susana and M. Àngels for their constant assistance and for always going above and beyond to ensure everything worked as it should. Within UB, I want to thank Aitana for all her work, and Jordi Roca, although words cannot do justice to his role. His administrative support has been key, but it goes far beyond “just” administrative matters. Thank you for your advice and for always taking care of me.

The department has also given me wonderful professors with whom to share insights and from whom to learn. Pilar, Judit, Jordi, Guillem, Jeni, Javier, and Asia: thank you for your continuous support and genuine interest over these years. I also want to thank José and José María for always keeping my best interests in mind; Rafa, for his invaluable insights on navigating the job market; and Antonio, for his interest and assistance on more than one occasion, sometimes even beyond working hours. Finally, I want to thank Miguel Ángel López and Inés Macho from UAB, and Pedro Rey from ESADE, who are the main culprits of my having started this journey in the first place. Your continued guidance, both before and after beginning my PhD, has been invaluable.

Since the day I began my PhD, I have been fortunate to meet amazing, hard-working colleagues, without whom these years would have been much more difficult. A special mention goes to Ghizlen, Claudia, Filippo, Marie, Ha, and Fernanda: your steps ahead served as a beacon to follow. And to those with whom I have walked side by side, especially Giovanni, Marianna, Mengwei, Sergio, and Juan: whether our time together was brief or plentiful, you have each helped me in one way or another to reach this point, and I will be forever grateful. I also want to mention my colleagues from afar: Adilya, Germán, Joan and Maria. Thank you for the talks, the advice, the memes, and all our moments together, sometimes difficult, sometimes joyful, but always shared. And to the people at Ivàlua: although we still have not spent much time together, I have thoroughly enjoyed it, and I have learned a lot from all of you.

To Nerea, Jordi, Manu and Víctor, my chosen family. Thank you for the laughs, the dinners, the lunches, and the coffees. Thank you for always being there, for being my refuge. And thank you to all the friends who have helped me along the way: whether in high school, at university, in Alicante, or in Lund. Our shared moments have all contributed, in one way or another, to this achievement.

Finally, some extraordinary people must be mentioned here: my family. Although I cannot name everyone, know that you have all shaped who I am today. To my parents especially, thank you for everything you have given me in life: for teaching me generosity and care for all living beings, for showing me never to give up, for providing me with the resources and opportunities to grow into who I am, often

at your own expense, and for standing by my side in the good times and especially in the bad ones. To my grandparents, Ana, Lola, Luis, and Juan, whose love, memories, and days spent together will always accompany me. And to my second family, my dogs: Guppy, Luna, and Fosca. Your unconditional love has no boundaries, and I have learned so much from you. Thinking of you always brings a smile, even on the hardest days, and fills my heart with your warmth.

Last but not least, I want to thank Jordi. I cannot express how grateful I am for having you in my life. Since day one, you have been a tremendous support, helping me navigate this maze. Thank you for extending your hand so we can walk this path together.

El meu sincer agraïment, de tot cor, a tots vosaltres.

Barcelona, September 30th, 2025.

Contents

1	Introduction	1
2	Building your Team: The Students Dilemma when Choosing Companion	5
2.1	Introduction	5
2.2	Background	8
2.3	Experimental design	9
2.4	Data and descriptive statistics	11
2.5	Results and Discussion	12
2.5.1	Students' preferences	12
2.5.2	Teachers' preferences	21
2.6	Summary and Concluding Remarks	26
2.7	Appendix	27
A1	Figures	27
A2	Tables	30
3	The Cost of Friendship: Who Is a Better Working Partner?	35
3.1	Introduction	35
3.2	Background	39
3.3	Experimental design	40
3.4	Data and descriptive statistics	42
3.5	Results	44
A1	Group composition effects on performance	44
A2	Well-being effects of the group	57
3.6	Summary and Concluding Remarks	64
3.7	Appendix	67
A1	Robustness checks	67
A2	Figures	69
A3	Tables	80
4	Employment Protection & Education Decisions	97
4.1	Introduction	97
4.2	Theoretical framework	99
4.3	Institutional framework	101
4.3.1	Employment at will	101
4.3.2	Exceptions of the employment at will doctrine	102
4.4	Data	102
4.5	Empirical Framework	107
4.6	Results and Discussion	110
4.6.1	Effects on the Labor Market	110

Contents

4.6.2	Effects on educational attainment	116
4.7	Summary and Concluding Remarks	121
4.8	Appendix	124
5	Concluding Remarks	135

1 Introduction

Human development and labor-market outcomes emerge from an interaction between social environments and the institutions that organize economic life. Early in life, pupils develop cognitive & non-cognitive skills inside peer networks and classrooms; later, firms and labor-market rules determine which skills and signals are rewarded. This dissertation brings those stages into conversation. Chapters 2 and 3 study how students form teams in school and how partner characteristics (ability, friendship, enmity, marginalization) affect both performance and well-being. Chapter 4 documents how changes in the costs of dismissing workers alter the returns to formal schooling and the distributional consequences of that re-pricing. Taken together, the three chapters map a plausible life-cycle pathway: micro decisions about who we work with at school affect what skills and social environments we are exposed to, and labor-market institutions determine how the observable outputs of schooling are valued.

Two sets of literature are at the core of this study. First, an extensive literature documents that peer composition and team processes matter for cognitive outcomes and for non-cognitive dimensions such as motivation and social skills (e.g. [Sacerdote \(2001\)](#); [Hoxby \(2000\)](#); [Weidmann and Deming \(2021\)](#)). Classroom and small-team settings create channels for tutoring, social reinforcement, and the development of collaborative competencies; they also create opportunities for exclusion or clustering that can amplify inequality in exposure to productive peers. Second, a separate literature shows that employers' information problems and institutional constraints (notably firing costs and employment protection) shape hiring behavior and the signaling value of credentials: when on-the-job adjustment is costly, firms rely more heavily on pre-hire signals such as formal education ([Spence \(1973\)](#)).

Both of them can affect each other. Classroom grouping processes determine the immediate social context in which cognitive and non-cognitive skills grow: they affect exposure to different types of peers, the quality and intensity of on-task learning, and students' comfort and morale, factors that can alter both measured achievement and longer-run preferences toward cooperative work. Later, hiring practices and dismissal rules determine which observable signals (such as formal education) employers rely on when selecting workers. If institutions make on-the-job adjustment costly, employers put more weight on pre-hiring signals and on observable credentials. Hence, the social processes that shape what students learn and whom they learn with interact with the institutional environment that ultimately prices that learning and the signals provided by that education.

This dissertation contributes to the existing literature in three different ways. First, by providing detailed, large-scale experimental evidence on the mechanisms behind team formation in class, how they are affected by characteristics, preferences and relations between the students. Second, by quantifying the causal effects of partner traits on both performance and well-being in real school settings, documenting a trade-off between comfort and efficiency that has practical implications for classroom design. And third, by showing that institutional changes in dismissal costs can materially re-price education, especially for younger workers, without necessarily inducing immediate additional schooling. Empirically, the work integrates experimental classroom evidence with population-level survey data and legal changes, thereby linking micro behavioral mechanisms to aggregate distributional outcomes.

The second chapter studies the determinants of team choice and the differences between student-chosen and teacher-assigned groups using a lab-in-the-field experiment in 19 schools in the Barcelona metropolitan area, with a final sample size of 1,735 students across 86 classroom groups.

The chapter's findings are situated in a rich set of empirical and theoretical contributions on how group composition and selection rules shape outcomes. Studies of endogenous versus exogenous grouping show that allowing choice increases homophily (such as gender) and can lead to polarization of group composition ([Aparicio Fenoll and Zaccagni \(2022\)](#)), while work on group decision-making and accountability argues that ties of familiarity or repeated interaction change incentives inside teams ([Thompson et al. \(1998\)](#); [Wang et al. \(2017\)](#)). By mapping both students' stated partner preferences and teachers' assignment logic, Chapter 2 shows how institutional rules (teacher assignment) mechanically alter exposure, for example, by deliberately mixing ability and integrating marginalized pupils, relative to student self-selection, which privileges comfort and perceived ability. These compositional shifts matter because they change the set of peer-tutoring and social-learning opportunities available to different students and therefore have predictable implications for short-run learning and the longer-run distribution of skills entering the labor market.

The data used in this chapter was obtained from two sources. On the one hand, students completed a survey with personal information as well as an individual math test used to classify their baseline ability level. Furthermore, they listed friendship and enmity relations, and stated preferred and undesired teammates. On the other hand, teachers provided their own hypothetical groupings.

Results found in this chapter show that students display strong homophily and performance considerations in partner selection: same-gender choices are common, higher-ability students are chosen more frequently, and socially marginalized students (classmates identified as bullied or of migrant origin) are substantially less likely to be selected as partners. Teachers, by contrast, form more ability-heterogeneous and inclusion-oriented groups and account for interpersonal relationships when assigning groups. Crucially, student-preferred and teacher-assigned partners diverge often: the partner a teacher assigns is not in the student's preference list in a majority of cases. These facts establish that the grouping mechanism (self-selection versus assignment) systematically changes group composition along dimensions that matter for both performance and welfare.

The third chapter exploits the same experimental platform but focuses on causal effects of pair composition by using random assignment of partners for the group testing phase, which happens after the baseline math test: individuals are randomly assigned a partner, and asked to complete a math test working together, as well as to answer some questions about their level of comfort when working together. Last, they are asked to solve a third, shorter individual math test.

The causal patterns documented here speak directly to multiple strands of the teamwork and peer-effect literatures. First, the asymmetric ability gains align with models of peer tutoring and comparative advantage: weaker students can benefit from collaboration (through explanations, scaffolding or increased engagement), while stronger students may lose due to coordination costs or effort substitution ([Goldman \(1965\)](#); [Webb \(1982\)](#)). Second, the contrasting effects of friendship and enmity connect to the social-incentives literature: [Bandiera et al. \(2010\)](#) and [Jehn and Shah \(1997\)](#) show that pre-existing ties alter effort incentives and conflict resolution inside teams; [Weidmann and Deming \(2021\)](#) emphasize social skills and "key contributors" as determinants of team success. Examining these patterns in a real school setting makes it possible to observe how students' characteristics and pre-existing social relations affect their performance and learning outcomes.

This chapter presents two main conclusions. First, ability composition matters asymmetrically: students with baseline ability below the median benefit from working in a pair: their group test scores rise for low-ability students regardless of their partner's ability; whereas students above the median do not experience gains and in some specifications see small declines in performance when paired with lower-ability partners. Quantitatively, a one-question increase in the student's baseline score raises group test performance by roughly 5.4% (and a one-question increase in the partner's baseline score raises the group score by approximately 4.8%), underscoring that both own and partner ability mechanically influence outcomes.

Second, social relations shape well-being but have ambiguous or negative effects on measured performance: being paired with a self-reported friend raises reported comfort by about 0.10 points (around 15% of the mean comfort score), while pairing with an "enemy"¹ strongly reduces comfort (0.27–0.37 points). Yet mutual friendship between pair members is associated with a decline in measured group performance (about a 5–6% drop, comparable in magnitude to a one-point lower baseline score), and mutual enmity causes substantially larger performance losses. Thus, friendship increases subjective well-being but can reduce short-run test performance; enmity hurts both comfort and performance. These results identify a clear trade-off between students' preference for comfort and short-run academic efficiency, in line with previous experimental work, such as that by Tudge (1992), suggesting that familiarity can increase off-task behavior or reduce accountability, producing a welfare trade-off between non-cognitive benefits and short-run cognitive efficiency. The randomized design isolates partner effects cleanly and quantifies magnitudes that are directly actionable for classroom policy (for example, the potential gain to low-ability students from pairing versus the cost to high-ability students and the welfare implications of persistent friend-based matching).

The fourth chapter studies how higher dismissal costs affect the wage return to schooling and whether individuals respond by extending schooling. To answer these questions, we exploit staggered, state-level judicial adoption of wrongful-discharge exceptions to the U.S. at-will doctrine (called implied-contract, public-policy and good-faith exceptions) between 1970 and 1999.

The analysis conducted in this chapter are grounded in the signaling and human-capital literatures (Becker (1964); Mincer (1974)) and in empirical work on the labor-market effects of employment protection (Autor et al. (2006); Dertouzos and Karoly (1993); Macleod and Nakavachara (2007)). The core mechanism emphasized is employer screening under costly post-hire adjustment: when dismissals are costly, employers rationally increase reliance on observable pre-hire signals (such as education) to reduce the risk of costly mismatches, a prediction consistent with the signaling theory (Spence (1973)) and with models of hiring under adverse selection (Kugler and Saint-Paul (2004)). The chapter also connects to empirical evidence that employment protection can both reduce flows and reshape job quality and on-the-job training (Autor et al. (2006), Cabrales et al. (2017)), which helps explain why an observed rise in the schooling premium need not immediately trigger extra schooling: credit or liquidity constraints, timing (many cohorts already passed prime schooling age), and the lagged nature of education investment can blunt the short-run supply response.

Using Census data and the Current Population Survey (CPS) and a differences-in-differences design that interacts education with policy exposure and age cohorts, the chapter finds that the adoption of dismissal-cost protections materially increases the wage premium to formal schooling, especially for younger cohorts (ages 15–19 at the time of reform), for whom the education premium under the policy is several times larger than for older workers.

¹As explained in more detail in the second chapter, we use this term to refer to students whom the focal student reports disliking.

Introduction

The implied-contract exception yields the largest and most robust premium effect, while the good-faith doctrine shows little effect, consistent with its sparse adoption. Importantly, despite clear increases in the market reward for schooling, there is virtually no detectable change in individuals' attained years of education after adoption: estimated effects on achieved schooling are centered near zero across specifications and data sources (Census and CPS). The pattern is consistent with a signaling interpretation: firms facing higher firing costs substitute toward workers with higher education levels, raising returns to schooling, and with this fact not inducing an observable supply response.

The chapter also documents modest distributional labor-market impacts: employment probabilities fall slightly for the least-educated and rise for more educated workers, again reinforcing the notion that higher dismissal costs tilt hiring toward more educated candidates and increase the premium for schooling without producing a measurable schooling increase within the cohorts and time window analyzed. The analysis for this chapter is complemented with several robustness checks, including varying control cohorts, using state-of-residence versus state-of-birth, and alternative outcome measures.

Finally, Chapter 5 provides concluding remarks, reviewing the main results of the previous chapters. Additionally, it addresses the main lessons drawn from this research in terms of policy design.

2 Building your Team: The Students Dilemma when Choosing Companion

2.1 Introduction

The process of team formation has been explored extensively in workplace environments, where factors such as individual skillsets, interpersonal dynamics, and organizational goals are critical in shaping the composition of teams. However, the study of its application in school settings remains comparatively underexplored, raising important questions about the mechanisms and implications of team formation among students. How do students select their peers for collaborative academic tasks, and what factors influence these decisions? Are preferences shaped by individual characteristics such as gender, academic ability, or pre-existing social connections like friendships? Is the decision of who to form a group with different from the decision of who to befriend? How are these decisions affected differently by individual characteristics? Moreover, do the criteria students prioritize when forming groups align with the strategies employed by teachers, or are there notable divergences between student-formed and teacher-formed teams? These questions are particularly relevant in classrooms, where collaborative projects have become an increasingly prevalent educational tool. Team work is often designed to achieve multiple objectives: fostering social skills, promoting teamwork, and facilitating peer-to-peer learning. Understanding the dynamics of team formation in schools is thus critical, as these decisions not only influence academic performance, but also impact classroom relationships and the social development of students.

As outlined in [Hill \(1982\)](#), group¹ interactions are often favored in a variety of contexts, and the process of group formation plays a critical role in determining both group performance and well-being during this process. Certain individuals, often referred to as “key contributors,” can significantly elevate the success of the groups they join [Weidmann and Deming \(2021\)](#), while the social connections between group members also influence outcomes. For instance, [Bandiera et al. \(2010\)](#) found that productivity varies depending on pre-existing relationships among coworkers. These findings highlight the importance of considering both individual traits and social dynamics when forming teams, especially in educational settings where group projects have become an integral component of students’ learning experiences throughout their academic journey, aiming to balance skill development with knowledge sharing.

Previous research also suggests that the composition of groups, including gender dynamics and familiarity among members, shapes the collaborative process in meaningful ways. Studies like [Thompson et al. \(1998\)](#) have shown that cohesive groups make more effective decisions, while [Hsieh et al. \(2020\)](#) revealed that interactions among strangers create unique dynamics that influence teamwork. Gender considerations are similarly important; [Leaper \(1998\)](#) demonstrated that responses to suggestions differ depending on the gender of group members, which can affect how decisions are reached. These findings indicate that the

¹In this paper we will refer interchangeably as “groups” or “teams” to the association of individuals working together on a task.

way groups are formed—whether based on compatibility, diversity, or random assignment—has a direct impact on their ability to achieve goals and on the experience of their members. Understanding how the group-formation process works can help optimize the decision on how to form them in both educational and professional contexts, leading to better outcomes depending on the specific objective of each situation.

This paper seeks to investigate the determinants of group formation in schools through the lens of a large-scale field experiment conducted across schools in the Barcelona metropolitan area. Our study utilizes detailed data on student characteristics, interpersonal relationships, and group preferences to explore the factors that shape decisions regarding group formation, both from the perspective of students and teachers. By comparing the composition of student-selected groups with those formed by teachers, we aim to provide new insights into the interplay between students and teachers decision-making processes. Specifically, we examine how gender, ability, social dynamics such as friendships, and other factors (e.g., bullying or migrant status) influence the formation of groups and how these factors vary depending on whether groups are self-formed or assigned by teachers. The role of group composition is especially critical given its potential implications for academic performance, peer learning, and well-being.

If group composition and preexisting relationships indeed affect student outcomes, as shown in [Izquierdo-Nofuentes \(2025\)](#), then the mechanism by which teams are formed may have far-reaching effects on students' academic trajectories. When students select their own groups, do they cluster in ways that optimize objective performance, or do they instead arrange themselves to maximize personal comfort and social cohesion? What will be the end result of this group-formation process? Conversely, when teachers assign groups, how do their priorities differ from students priorities? Do they prioritize group-formation based on performance, or do they place greater weight on the well-being and interpersonal dynamics of participants? Will the resulting classroom organization differ when students form their own groups compared to teacher-assigned teams? Our findings shed light on these questions, with significant implications for educational policy and practice.

This paper contributes to the literature by examining group formation in a real-world educational setting, with a focus on gender, ability, and social dynamics such as bullying, migration status and friendship. Additionally, we assess how teacher-formed groups differ from student-formed ones, offering insights into whether teacher interventions align with or diverge from student preferences and, therefore, which would be the optimal group-formation mechanism to use, depending on the purpose we may have in mind. To the best of our knowledge, this study is the first to investigate the factors influencing group formation in schools and the differences between student- and teacher-formed groups. While the effects of specific characteristics on group composition have been explored in related contexts, our study adds new dimensions to this field of research. For instance, [Aparicio Fenoll and Zaccagni \(2022\)](#) analyzed how allowing students to choose their partners affects group gender composition, finding that self-formed groups tend to be more gender-polarized compared to randomly formed ones. This aligns with our findings, which show that gender plays a key role in students' preferences for group formation.

Furthermore, [de Matos Fernandes et al. \(2024\)](#) explored the decision-making process students use when forming teams, focusing on whether these choices are driven by shared characteristics or existing relationships. Our findings build upon this work by showing that friendship and shared characteristics strongly influence students' partner preferences. In a complementary line of inquiry, [Wang et al. \(2017\)](#) analyzed how group decision-making differs from individual decision-making in risky contexts, emphasizing that group decisions inherently involve accountability to teammates, which may, in turn, impact how students decide whom to form a team with, knowing that it may also influence the level of effort their partners may put into collaborative school tasks.

In this study, we implemented a field experiment involving 1,735 students from schools in the metropolitan area of Barcelona. The experimental sessions lasted approximately two hours and were structured to include several components. First, participants completed a basic information survey, followed by the construction of a student relationship network and a series of questions about their preferred teammates to work in group for a given class assignment. The survey also gathered information on students' suffering from bullying, since this can be an influencing factor when deciding whom to team up with. Additionally, teachers were asked to complete a brief survey detailing how they would form student groups for classroom assignments. This survey was conducted as part of a larger study in which, after completing the survey, students had to answer additional questions and participate in different math tests, some of those individually and some in groups, to analyze the effects of group work on performance and well-being. The results of this study are reported in [Izquierdo-Nofuentes \(2025\)](#).

The sample predominantly comprised secondary school students aged 12 to 16, along with students in the final two years of elementary school (ages 10 to 12) and those enrolled in vocational training programs (aged 16 and above). These students were recruited from both public and charter schools across various cities and neighborhoods with distinct socioeconomic characteristics.

Our findings reveal that students exhibit strong preferences for working with peers of the same gender and those with higher academic ability, while often avoiding classmates who are socially marginalized (i.e. those experiencing bullying or with a different origin). Gender homophily is particularly pronounced: both male and female students disproportionately select same-gender partners. High-ability students are more likely to be chosen, reflecting perceptions of their capacity to boost group performance.² Conversely, bullied and migrant students face significant obstacles to selection. In contrast, teachers form more heterogeneous teams, deliberately balancing gender, ability levels, and social inclusion. Comparing student-formed and teacher-formed groups highlights how individual preferences can diverge from institutional objectives. The choice of grouping mechanism (student self-selection versus teacher assignment) may translate into different compositions which, in turn, may influence students outcomes, including academic performance, well-being, and the inclusion of marginalized students.

Beyond team formation, we also analyze the aspects behind friendship formation, revealing both similarities and differences. Students select their friends using similar criteria as when choosing teammates, with ability, gender, migrant status, and bullying status all playing a role. However, the magnitude of these effects is smaller in friendship formation compared to team formation. For example, while ability plays a positive role in both cases, the impact on friendship selection is less pronounced, likely because friendships develop over longer periods and involve repeated interactions, whereas team selections are short-term and task-focused.³

One notable difference emerges in the role of gender: gender has no significant effect on the overall probability of being selected for a group, but it does affect who selects you, as male students are more likely to be selected by other male students and the same applies to female students. However, being female does negatively affect the likelihood of being chosen as a friend. On average, female students are selected as

²We also find that students who overestimate their own math ability are more frequently selected, suggesting that perceived ability influences team selection independently of actual ability.

³Note that when students were asked to choose classmates for group work, no particular type of task was specified. The prompt simply read: *Select the classmates with whom you would like to form a group to work on a classroom task.*

friends less frequently than male students. This pattern stems from two factors. First, both male and female students display strong gender homophily, consistently preferring same-gender peers, as observed also in group selection. Second, female students nominate fewer friends overall than male students.⁴ Together, these two dynamics contribute to the observed negative impact of being female on friendship selection.

Teachers, in contrast, adopt a more structured and integrative approach when forming groups. While their choices also reflect gender considerations, they show less emphasis on gender homophily than students. Importantly, teachers tend to prioritize diversity in ability by pairing students with different academic skill levels, such as high-ability students with low-ability peers. This strategy contrasts sharply with the student-formed groups, which would likely cluster high-ability students together due to mutual preferences. For socially marginalized students, teachers avoid pairing individuals who are both victims of bullying, opting instead to integrate them with non-bullied peers. Similarly, migrant students are predominantly paired with local students, fostering inclusion and avoiding isolation within the group dynamic.

Furthermore, we identify differences in how friendships influence group formation. While teachers account for friendship networks in their group allocation, this result is similar to the one found if groups were randomly formed. Nonetheless, teachers make improvements by effectively accounting for negative relationships: they form fewer groups consisting of students displaying enmity and tend to group students with neutral relationships more frequently.

On the other hand, social preferences, as captured by our dictator game indicators, appear to have no significant influence on team formation, whether in student-formed groups or teacher-formed groups. This absence of impact suggests that social preferences are overlooked during the group formation process.

When comparing teacher-formed groups to those that could have emerged based on student preferences, we find a mismatch. While teachers rarely assign students to groups with their explicitly undesired peers, they also frequently overlook preferred pairings. This divergence suggests that teachers, while aware of the interpersonal relationships within the classroom, prioritize factors such as ability diversity and social integration over individual preferences.

The remainder of the paper is organized as follows. [Section 2](#) provides background information about the experiment. [Section 3](#) describes the experimental design and implementation. [Section 4](#) presents the data collected. In [Section 5](#), we present the results of our analysis. Finally, [Section 6](#) summarizes and concludes.

2.2 Background

Prior to conducting the experiment, we preregistered our primary hypotheses and methodological approach with the [Registry of Efficacy and Effectiveness Studies](#). Additionally, conducting research with minors requires adherence to strict protections and laws, which is why we sought and received external ethical approval from the [Bioethics Commission](#) at the University of Barcelona,⁵ which formally approved the study on November 2nd, 2022. Furthermore, all responses were guaranteed to be anonymized, ensuring that researchers are unable to identify the participants. To address challenges associated with non-standard samples and to reduce the incidence of missing data, we streamlined the format of responses

⁴This difference was not evident in group formation due to the imposed upper limit of five groupmate selections, whereas the friendship question had no such constraint.

⁵Institutional Review Board IRB00003099.

by primarily employing multiple-choice questions in lieu of open-ended ones. The software's design made it compulsory for participants to respond to all questions before proceeding to the next one, thereby preventing the omission of answers. Nevertheless, for subjects that might be perceived as sensitive, participants were offered the option to select "I would prefer not to answer."

We conducted experimental sessions in 19 schools located in the Metropolitan Area of Barcelona. Initially, we reached out via email to all schools receiving funding from the Education Department of the Generalitat de Catalunya to identify those potentially interested in participating. After explaining the objectives of our research and meeting with several school representatives, we finalized the list of schools that would comprise our sample. The participant pool was recruited through agreements with school headmasters, who consented to conduct the experimental session as a classroom activity integrated into their pedagogical curriculum, which resulted in a high level of participation. The experimental sessions were conducted on-site at the schools, using an online software named Social Analysis and Network Data ([SAND](#)),⁶ which assisted in mitigating privacy concerns. Students could use any of their devices (laptop, classroom computer, tablet or smartphone) in order to access and carry out the experiment. The questionnaire was administered entirely in Catalan. Prior to the experimental sessions, the schools provided us with signed consent forms, agreeing to conduct the session as an official school activity, along with class lists detailing the enrolled students. Using these lists, we created unique usernames and passwords for each student in the software employed for the experiment. We then randomly paired students using randomization software, forming the groups that would collaborate during the session. For classes with an odd number of students, the remaining student was designated as a "wildcard" to be assigned on the session day if any pre-assigned partner was absent.⁷

2.3 Experimental design

The experiment is structured into three distinct phases, as outlined in [Fig. 2.1](#). During the first phase, students complete a survey designed to gather comprehensive information on their sociodemographics, family background, self-assessed mathematical competence, and academic performance in the previous years. Subsequently, participants are prompted to indicate the level of friendship⁸ they share with their classmates and to select individuals they would like and those they would like not to collaborate with on a school task. Screenshots of the questions pertaining to students' social networks are provided in [Fig. A2.1](#) and [Fig. A2.2](#) in the Appendix. This information enables us to analyze decision-making processes regarding friendship and partnership selection, the characteristics prioritized by students in forming these relation-

⁶This software was commissioned by an amalgamation of public universities and institutions with the purpose of obtaining data from teenagers in schools across Spain. For more information we refer the reader to [Vasco et al. \(2025\)](#).

⁷The number of students who ended up working alone was very small (36 students, 2% of our total sample).

⁸Participants were first instructed to identify individuals in their class whom they regard as friends. From this group, they were then asked to select those they consider their best friends. Subsequently, participants were instructed to indicate individuals with whom they do not maintain a positive relationship (for the purposes of this paper, these individuals are referred to as "enemies" consistent with the terminology used during the experimental session). From this list, they were then asked to select those with whom they have the most difficult relationships.

ships, and the possible differences when selecting a friend instead of your preferences about your group partners. At the end of the first phase, students are required to complete a series of timed multiple-choice tests in Mathematics⁹ individually, which will be used as a baseline to compare with the subsequent tests.¹⁰

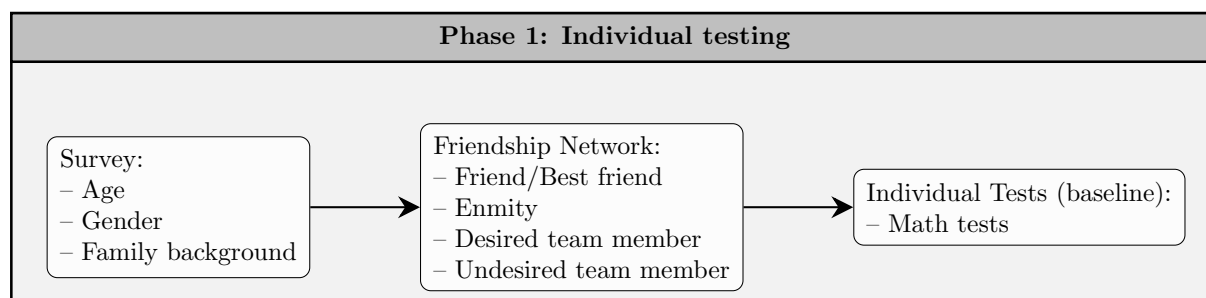


Fig. 2.1. Chronology of the experimental sessions

The second phase commences with the random assignment of each student to a peer, with the paired students tasked to jointly undertake another set of multiple-choice tests, in which they are allowed to discuss how to solve the problem to come up with the potential correct solution. Upon completion, participants are separated and asked regarding their comfort levels when working with their specifically assigned partner.

In the final phase, students individually complete a third battery of multiple-choice tests. Additionally, they are requested to answer some questions about their morale state and to identify any classmates, including themselves, who may be experiencing bullying. With this information, we are able to explore the relationship between performance in the different tests and the individual characteristics of the student and their partner.

Simultaneously, data regarding teachers' preferences for group formation is collected. Specifically, teachers are asked to create hypothetical groups for a classroom task, as well as the reason behind forming these specific groups.

During the experimental session, students are organized by grade and class and seated according to their usual classroom arrangement. A set of instructions, presented in Catalan, is displayed on their screens and read aloud to the students: "Dear students, we are carrying out a lab-in-the-field experiment in the field to understand how social relationships affect group performance. Please, work in silence and answer the questions truthfully." Subsequently, students receive guidance on how to access the experimental platform and the procedure for identifying their randomly assigned partner in the group phase.

Students do not receive feedback on their performance at any point during the session. As explained before, each participant is assigned a unique username and password, which enables them to respond independently on their device, particularly if they disagree with their partner during the group phase. Additionally, we record the time at which each student answers each question. This timing data allows us to monitor instances of students potentially selecting random answers and to identify whether participants are collaborating simultaneously with their assigned partners.

⁹Mathematics tests questions were classified according to the grade of the student. For elementary school students, we use questions extracted from [Proves Cangur](#) corresponding to the US-equivalent 5th and 6th grades. For the remaining students, we use questions based from the [ASVAB](#) test, and classify them by grade according to school syllabi.

¹⁰For a detailed analysis of the results regarding the performance of students on the tests, readers are referred to [Izquierdo-Nofuentes \(2025\)](#).

2.4 Data and descriptive statistics

Our sample is composed of students from nineteen educational centers, including elementary schools, secondary schools, and vocational training schools, located in the vicinity of Barcelona, with a total of 1,945 potential participants. All public and charter schools in the area were contacted to offer the possibility of participating. The final sample consists of 86 classroom groups and 1,735 students, following the exclusion of those absent on the day of the experiment or who were unable to be paired.

Each experimental session lasted for up to two consecutive hours, although students were permitted to finish earlier. No breaks were allowed during the sessions, and responses were recorded up until the point where each student completed the last question, if time constraints prevented them from finishing all of them. Most participants completed at least the first two tests, and many also completed the final test. This information is synthesized in [Table 3.1](#) below.¹¹

Table 2.1: Student distribution across tests by gender, ability, and bullying status

Test Phase	Male	Female	High-performing	Low-performing	Bullying	No Bullying	Total
Baseline Test	848	851	728	971	254	1445	1699
Group Test	847	850	726	971	253	1444	1697
Final Test	825	831	707	949	249	1407	1656

[Table 3.2](#) presents the main summary statistics for the primary sociodemographic characteristics of our sample. As previously detailed, the sample consists of students from various educational levels: elementary school students comprise 27% of the sample, secondary school students comprise 53%, and vocational school students comprise 20%. The gender distribution is balanced, with approximately equal numbers of male and female students, and the average age is 14.65 years. Additionally, 15% of the students have been identified by their classmates as experiencing bullying, are 7% of migrant origin.

The characteristics of the schools in our sample are detailed in [Table A3.1](#) in the Appendix. The sample includes 8 charter schools and 11 public schools. 10 schools are situated within Barcelona city, while the remaining nine are located in neighboring municipalities. Of these, four schools are located in areas with a socioeconomic index below the Catalonia average, eight have an index above but close to the average, and the remaining seven have an index well above the average. On average, the schools in our dataset have a socioeconomic index that is 7% higher than the Catalonia average.

¹¹For this analysis, we use exclusively the data from the baseline test, using the students' performance on this test to categorize them by ability level, depending on whether they are above or below the median value.

Table 2.2: Sociodemographics

	Mean	Median	Std. Dev.	Min.	Max.	N
Age	14.65	14.00	3.26	11.00	46.00	1735
Female	0.51	1.00	0.50	0.00	1.00	1735
Bullying status	0.15	0.00	0.36	0.00	1.00	1735
Migrant status	0.07	0.00	0.25	0.00	1.00	1676
Elementary school (10-12 years old)	0.27	0.00	0.44	0.00	1.00	1735
Secondary school (12-16 years old)	0.53	1.00	0.50	0.00	1.00	1735
Vocational school (16-46 years old)	0.20	0.00	0.40	0.00	1.00	1735

2.5 Results and Discussion

2.5.1 Students' preferences

Team formation

We begin by investigating how the individual characteristics of students influence their decisions when selecting teammates for a school project. Using a cross-sectional analysis, we focus on the sociodemographic attributes of the students being chosen, and the interaction with the attributes of the student making the selection. This dual approach allows us to identify patterns and preferences in group formation based on specific individual traits. In particular, we examine the roles of four key indicators: gender, ability level, migrant status, and bullying status, to understand the preferences shaping group selection.

To begin this analysis, we focus solely on the characteristics of the individual being chosen, to examine their effect on the likelihood of selection. We perform an Ordinary Least Squares (OLS) regression analysis of the probability of being selected to be on a team based on the individual characteristics of the student, employing the following equation:

$$\begin{aligned}
 Prob. Being Selected_i = & \beta_0 + \beta_1 Female_i + \beta_2 Ability_i \\
 & + \beta_3 Migrant_i + \beta_4 Bullying_i + \alpha_g + \epsilon_i
 \end{aligned}
 \tag{2.1}$$

A visual summary of these findings is provided in [Figure 2.2](#), where the dashed line represents the average probability of being selected at random across the sample, serving as a baseline for comparison. Detailed regression outputs corresponding to this analysis are presented in [Table 3.3](#). Our results indicate a significant positive effect of ability on the likelihood of being selected: students in a higher ability quartile

are approximately 9.7% more likely to be chosen compared to those in the quartile below. Conversely, we observe strong negative effects for both migrant status and bullying status. Migrant students are about 25% less likely to be selected compared to their local peers, while students identified as suffering from bullying are chosen roughly 35% less often than their classmates.¹²

Interestingly, we find no overall effect of gender on the probability of being selected to join a team, which prompted us to investigate further into the potential role of gender interactions in the decision-making process.

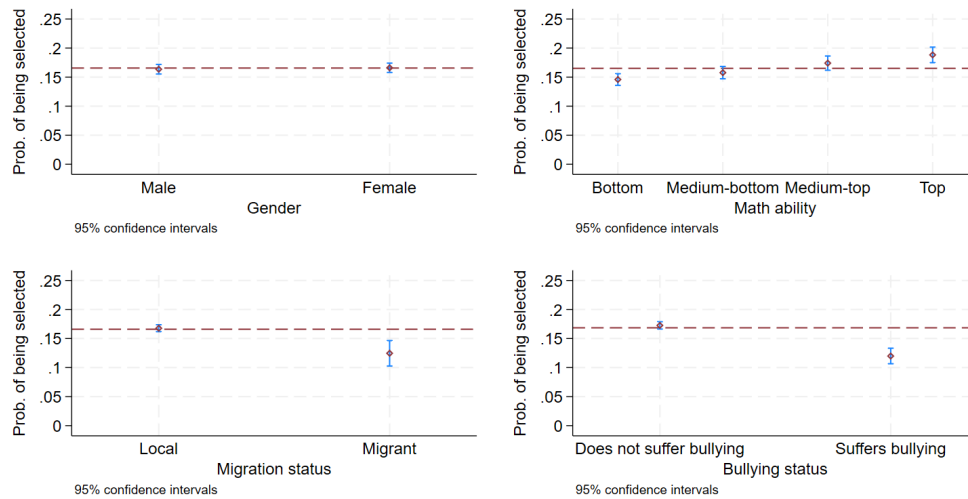


Fig. 2.2. Prob. of being selected depending on characteristics of the student selected

Table 2.3: Regression of individual characteristics on probability of being chosen for a team

	(1)	(2)	(3)	(4)	(5)
Female	0.00396 (0.00663)				0.00359 (0.00691)
Ability		0.0173*** (0.00279)			0.0160*** (0.00263)
Migrant status			-0.0417*** (0.0133)		-0.0409*** (0.0133)
Bullying status				-0.0596*** (0.00756)	-0.0568*** (0.00736)
Class FE	✓	✓	✓	✓	✓
Observations	1676	1676	1676	1676	1676
Mean of Dep. Variable	0.165	0.165	0.165	0.165	0.165

Note: Standard errors in parentheses, clustered by class. $p < 0.01$, $p < 0.05$, $p < 0.10$

¹²To address potential concerns that migrant students may be disproportionately affected by bullying, which could confound these results, we examined the correlation between migration and bullying status. We found the correlation to be extremely small and slightly negative, indicating that these variables operate largely independently.

In Figure 2.3, we explore the differences in the probability of being selected for a team based on both the gender of the student selecting and the gender of the student being selected. The corresponding detailed regression results are displayed in Table 3.4. Our analysis reveals a striking homophily effect, where students exhibit a strong preference for selecting peers of the same gender. Specifically, we find that the probability of being chosen by someone of your own gender is approximately 80% higher than that of being chosen by someone of the opposite gender. This pronounced preference underscores the significant role of gender similarity in team formation.

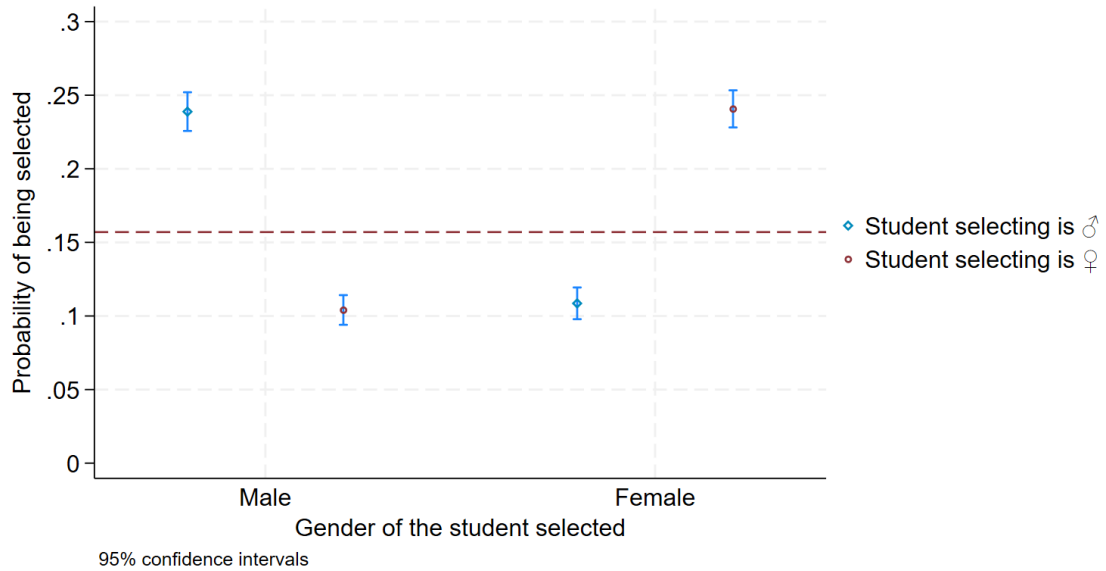


Fig. 2.3. Prob. of being selected depending on gender of the student selected and the student selecting

Table 2.4: Regression of individual characteristics on the probability of being chosen for a team

	(1) Prob. of being selected by a male student	(2) Prob. of being selected by a female student
Female	-0.136*** (0.0130)	0.146*** (0.0128)
Ability	0.0200*** (0.00437)	0.0142*** (0.00395)
Migrant status	-0.0349* (0.0199)	-0.0357** (0.0167)
Bullying status	-0.0664*** (0.0123)	-0.0600*** (0.0113)
Class FE	✓	✓
Observations	1672	1676
Mean of Dep. Variable	0.173	0.174

Note: Standard errors in parentheses, clustered by class. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Interestingly, the other variables under consideration —ability, migrant status, and bullying status— display consistent effects across both male and female selectors. This suggests that both male and female students evaluate these characteristics in a comparable manner when deciding on their teammates, further reinforcing the broader patterns identified in our general analysis.¹³

The last analysis of this section incorporates two distinct indicators: social preferences¹⁴ and self-perceived math ability¹⁵. Figure 2.4 offers a visual representation of these findings. Regarding social preferences, the results indicate no consistent relationship with the probability of being selected as a team member.¹⁶ On the other hand, the analysis of perceived math ability reveals a notable pattern similar to the one observed for actual math ability: students who perceive themselves to have higher math ability are more frequently chosen by their peers to join their teams. This finding suggests that students' subjective beliefs about their skills play an important role in group formation.

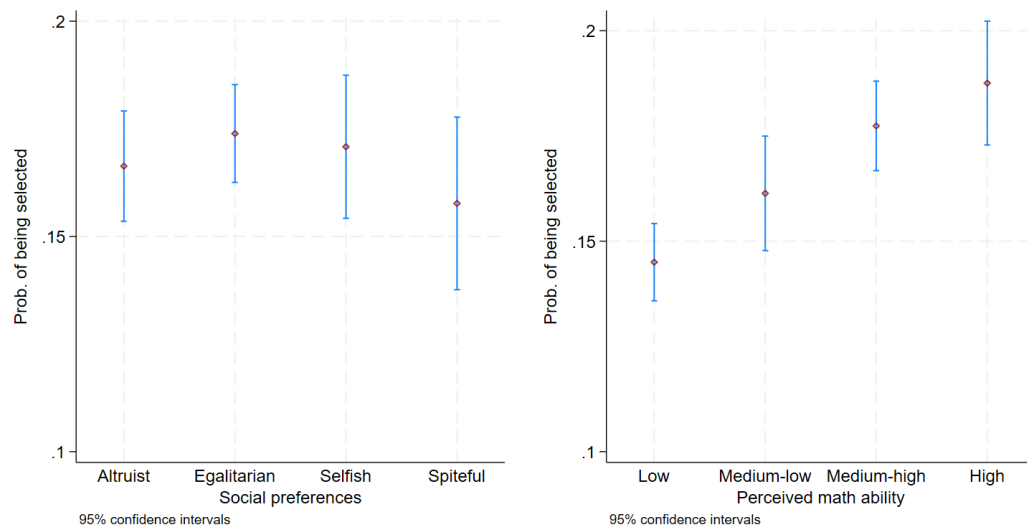


Fig. 2.4. Prob. of being selected depending on characteristics of the student selected

Specifically, as detailed in Table 3.5, a one-quartile increase in perceived math ability is associated with an 11.1% higher likelihood of being chosen to join a team. This result aligns closely with the impact observed for actual math ability, where the effect was 9.7%. In the final column of Table 3.5, we include both perceived and actual math abilities simultaneously to account for their potential correlation. Both variables remain statistically significant and positively associated with the probability of being selected, though their magnitudes are slightly reduced to 9.41% and 8.23%, respectively. These findings suggest

¹³A visual representation of the results depending on gender for ability, migrant and bullying status is shown in the appendix, in Figure A2.3 - Figure A2.5.

¹⁴Social preferences were measured using three variations of the dictator game, administered in random order. Based on their responses, students were classified according to their social preference type: altruist if they chose to benefit the other, egalitarian if they always chose the option to divide the money equally, selfish if they chose to take the money from the other player and spiteful if they chose to burn the money of the other player. Screenshots of the questions can be found in Figure A2.6 - Figure A2.8 in the appendix.

¹⁵Students were asked to evaluate their proficiency in three specific aspects of mathematics. This information was used to construct a composite indicator. While it is moderately correlated with actual math ability (correlation coefficient of 25%), we treat it as a separate dimension for analysis. Detailed results about the determinants of students perceived ability are shown in Table A3.2 in the Appendix.

¹⁶When we interact this variable with gender, we still find no significant differences, suggesting that neither male nor female students factor this characteristic into their team selection decision. Detailed results are shown in Table A3.3 in the Appendix.

that students' subjective beliefs about their skills play an important role in group formation. Specifically, when students decide whom to form a group with, they likely consider an amalgamation of their peer's actual math ability and their perception of how proficient that peer is. ¹⁷

Table 2.5: Regression of individual characteristics on the probability of being chosen for a team

	(1)	(2)	(3)	(4)
Social preferences	-0.000329 (0.00341)		0.00129 (0.00341)	0.000720 (0.00346)
Perceived math ability		0.0165*** (0.00253)	0.0189*** (0.00328)	0.0160*** (0.00334)
Female				0.00645 (0.00832)
Ability				0.0140*** (0.00340)
Migrant status				-0.0493*** (0.0172)
Bullying status				-0.0590*** (0.00853)
Class FE	✓	✓	✓	✓
Observations	1134	1656	1118	1118
Mean of Dep. Variable	0.169	0.165	0.170	0.170

Note: Standard errors in parentheses, clustered by class. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

¹⁷When we include the interaction between this variable and gender, the interaction coefficient is negative and significant at the 10% level. However, its magnitude is smaller than that of perceived ability, suggesting that although both male and female students factor perceived ability into their partner choices, this consideration is somewhat stronger among male students. This pattern mirrors our findings on actual ability, where male students placed greater emphasis on ability when selecting teammates than did female students. Detailed results are shown in [Table A3.3](#) in the Appendix.

To explore this relationship further, we focus on students who perceive themselves as strong in math,¹⁸ as reflected in their responses to the belief-based questions, but whose baseline test results indicate otherwise. The impact of being overconfident in one's math ability on the likelihood of being chosen for a team is analyzed in Table 3.6. Our findings reveal that students who incorrectly believe they are good at math are chosen more frequently than their peers. Specifically, this overconfidence translates into a positive impact of approximately 16% on the probability of being selected, even after accounting for other student characteristics in the model, including actual ability.¹⁹

These results suggest that the perception of ability plays a significant role in team selection, potentially outweighing actual ability itself. This highlights the importance of how students' self-assessments influence their desirability as teammates, emphasizing the social value of confidence or perceived competence in the group formation process and thus, the importance of strengthening students' self-confidence in their abilities, particularly among those with lower measured ability.

Table 2.6: Regression of math perception effect on the prob. of being chosen

	(1)	(2)
Bad Student Believed Good	0.0219** (0.00942)	0.0258** (0.0114)
Class FE	✓	✓
Controls	×	✓
Observations	919	621
Mean of Dep. Variable	0.152	0.158

Standard errors in parentheses, clustered by class. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Friendship formation

In the next stage of our analysis, we examine the formation of friendships among students and compare these patterns with the criteria students use to select teammates for school projects. Because students typically select at least one friend as a teammate, we proceed to examine the differences and similarities between these two choices. We include the same characteristics as in the previous section: gender, ability, migration status, and bullying status, and employ a similar empirical specification as in the previous analysis.

¹⁸To better understand the factors associated with students' self-perceived math ability, we conduct a regression analysis of its main predictors. The detailed results are presented in Table A3.4, showing that, while experiencing bullying and being of migrant origin have no effect on self-perceived ability, actual ability has a positive effect, whereas being a female student has a strong negative effect on self-perceived ability, consistent with previous literature and equivalent to a three-quartile decrease in measured ability.

¹⁹Another potential explanation for the effect of perceived ability is the popularity of the student. To investigate this, we estimate a regression model on the probability of being chosen for a team, including both perceived math ability and student popularity as explanatory variables. We use two distinct measures of popularity: the number of times a student is selected as a friend by peers and the number of friends the student personally selects. The detailed results, presented in Table A3.5, indicate that perceived math ability remains a statistically significant predictor of selection, even after controlling for student popularity.

$$\begin{aligned} \text{Prob. Being Selected as Friend}_i = & \beta_0 + \beta_1 \text{Female}_i + \beta_2 \text{Ability}_i \\ & + \beta_3 \text{Migrant}_i + \beta_4 \text{Bullying}_i + \alpha_g + \varepsilon_i \end{aligned} \quad (2.2)$$

Our objective is to explore whether the criteria students use to form friendships align with or diverge from the factors they consider when forming teams for school projects. This approach sheds light on whether group formation for school projects reflects broader social preferences or follows a distinct logic. A visual representation of these findings can be seen in [Figure 2.5](#), while detailed results are available in [Table 3.7](#).

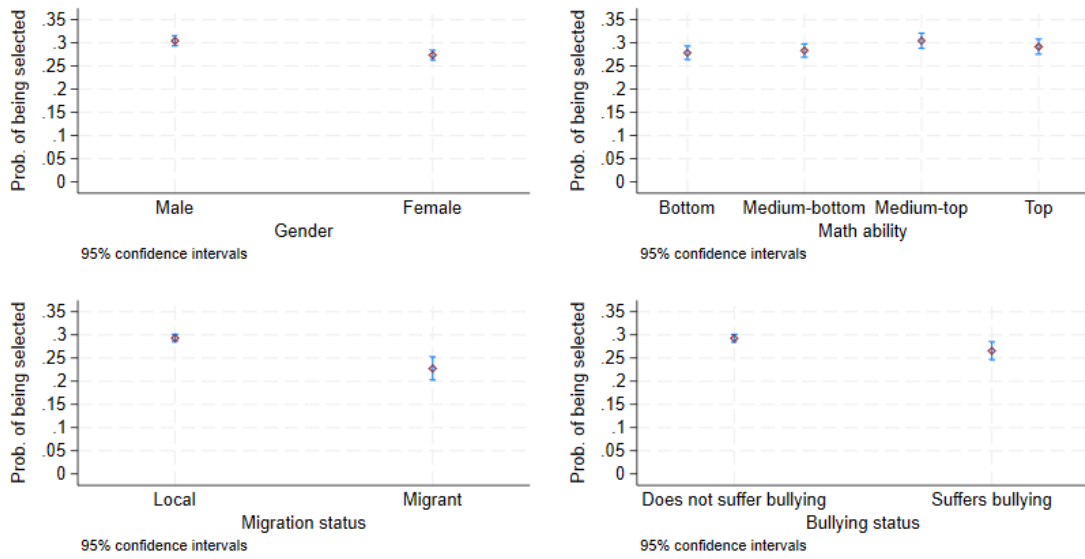


Fig. 2.5. Prob. of being selected as a friend depending on the characteristics of the student selected

When students decide whom to be friends with, all of the examined characteristics show an impact. However, the magnitude of these effects is noticeably smaller than in the context of team selection. Specifically, higher ability students are 2.64% more likely to be selected as friends compared to their peers in the lower quartile. Additionally, migrant students and those experiencing bullying have a negative impact on their probability of being selected as friends, with reductions of roughly 20% and 18%, respectively. These figures are considerably smaller than the 25% and 35% reductions observed in the team selection process. Interestingly, the relative importance of migrant status and bullying status has reversed in this context: while bullying had a greater negative impact on team selection, migrant status now exerts a stronger influence on friendship formation.

One potential explanation for these patterns is the temporal difference between the two types of social interactions. Friendships tend to be long-term relationships, whereas groupings for class projects are often short-term. When a new migrant student enters a classroom, friendship networks may already be established, and thus, the new student may face greater challenges in forming friendships.

Another notable difference between team formation and friendship selection arises with respect to gender. While being female had no impact on the overall probability of being chosen as a team member (only affecting who selected you), it has a significant negative effect of 6.12% on the overall probability of being selected as a friend. To delve deeper into this result, we conduct two additional analyses.

Table 2.7: Regression of individual characteristics on probability of being chosen as a friend

	(1)	(2)	(3)	(4)	(5)
Female	-0.0177*** (0.00655)				-0.0202*** (0.00687)
Ability		0.00942*** (0.00272)			0.00677*** (0.00247)
Migrant status			-0.0579*** (0.0144)		-0.0594*** (0.0144)
Bullying status				-0.0512*** (0.00742)	-0.0534*** (0.00756)
Class FE	✓	✓	✓	✓	✓
Observations	1676	1676	1676	1676	1676
Mean of Dep. Variable	0.289	0.289	0.289	0.289	0.289

Standard errors in parentheses, clustered by course.

$p < 0.10$, $p < 0.05$, $p < 0.01$

First, we revisited gender patterns in friendship selection, mirroring our earlier approach to team formation. These results are displayed in [Figure 2.6](#), where we observe similar trends as those found in the team selection process. Male students predominantly select other male students as friends, while female students exhibit a preference for selecting other female students. Detailed results are reported in [Table 3.8](#), which shows that the magnitude of this same-gender preference is approximately 85%.

Second, we analyzed the average number of friends selected by male and female students. In the team selection scenario, students were restricted to choosing a maximum of five peers; however, the friendship selection process allowed for an unrestricted number of choices. If male and female students differ in the number of friends they typically select, this could help explain the gender effect observed in our earlier analysis. These findings are presented in [Figure 2.7](#).

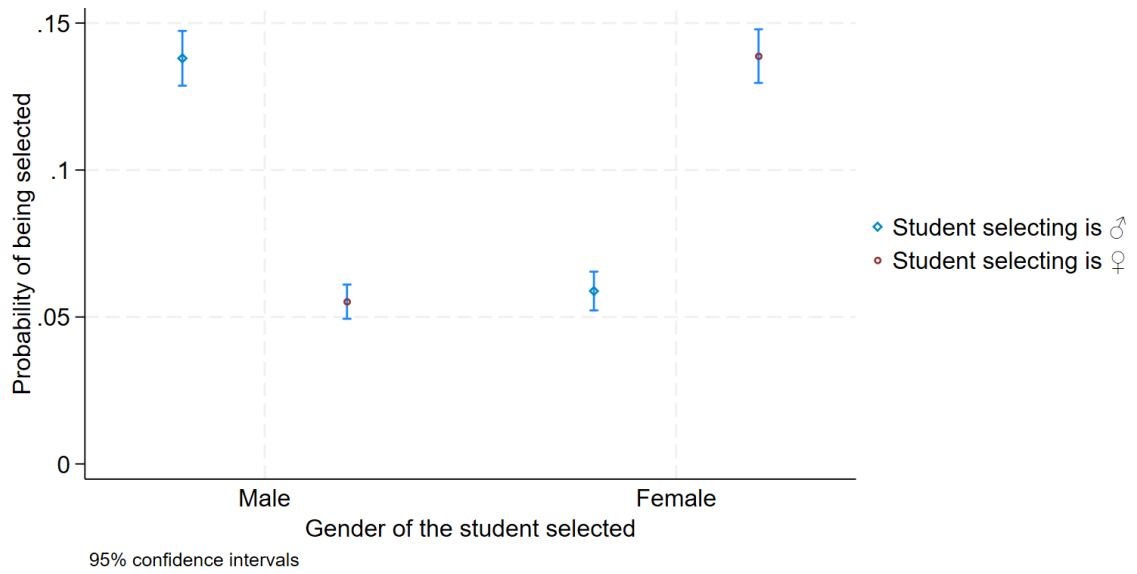


Fig. 2.6. Prob. of being selected as a friend depending on the gender of the student selected and gender of the student selecting

Table 2.8: Regression of individual characteristics on the probability of being chosen as a friend

	(1) Prob. of being selected by a male friend	(2) Prob. of being selected by a female friend
Female	-0.0821*** (0.0111)	0.0896*** (0.0112)
Ability	0.0118*** (0.00319)	0.0110*** (0.00285)
Migrant status	-0.0245*** (0.00907)	-0.0241** (0.00947)
Bullying status	-0.0346*** (0.0107)	-0.0346*** (0.00725)
Class FE	✓	✓
Observations	1676	1676
Mean of Dep. Variable	0.0977	0.0977

Standard errors in parentheses, clustered by course. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Our analysis reveals that male students, on average, select more friends than female students, with this difference being statistically significant at conventional levels. Specifically, male students report an average of 14.85 friends, compared to 12.35 for female students. This discrepancy may reflect differences in the actual number of friends or stricter criteria employed by female students in designating someone as a friend. This variation in the number of friends selected likely contributes to the negative effect of being

female observed in the probability of being chosen as a friend: given that female students are selected mostly by other female students, and that female students select fewer friends than males on average, this helps explain why being female has a negative impact on the overall probability of being chosen as a friend.

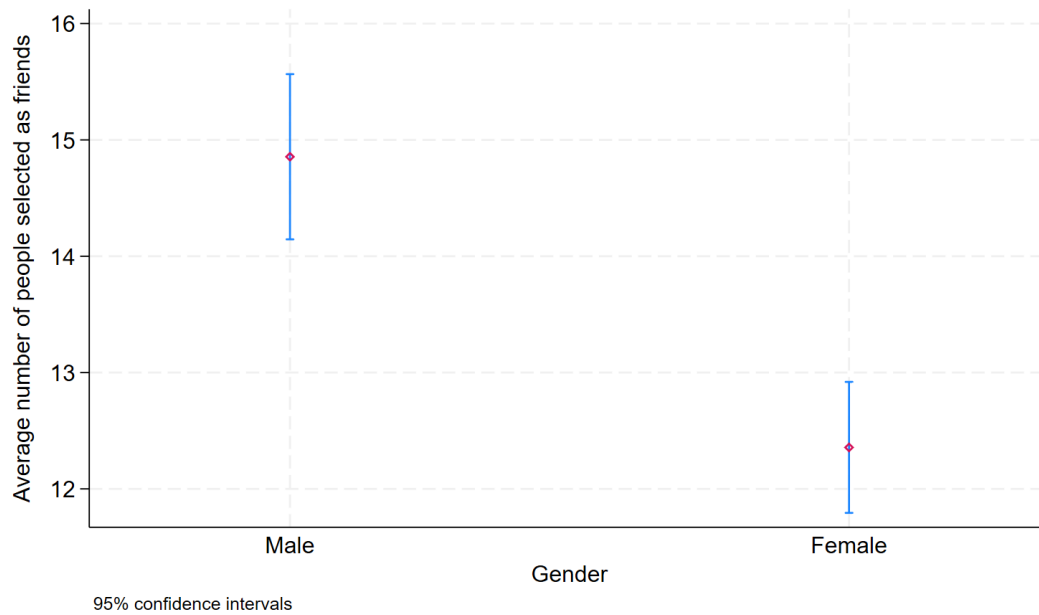


Fig. 2.7. Average number of friends selected depending on gender of the student selecting

2.5.2 Teachers' preferences

We now turn to the analysis of how teachers incorporate students' characteristics when forming groups for collaborative tasks in class, focusing on the four key sociodemographic indicators used in our previous analysis: gender, ability, migrant status, and bullying status. By examining the decision-making processes of teachers, we can better understand whether their methods align with or diverge from the preferences observed in student-formed teams. Since, in this case, we are dealing with actual group allocations rather than preferences, we analyze the characteristics of both members within each group. The results of this analysis are presented in [Figure 2.8](#).

Starting with gender, we observe that teachers tend to form more same-sex groups than mixed-sex groups, although the homophily component that we observed in student-selected groups is much less pronounced here. The numbers of same-sex and mixed-sex groups are more balanced compared to students' preferences.

Regarding ability, teachers typically pair high-ability students with low-ability students, creating mixed-ability groups. This differs from the student selection process, where high-ability students are more likely to be selected by their peers, and if students were allowed to choose, high-ability students would likely end up working with others of similar ability. In contrast, teachers opt for a more balanced approach, pairing high-ability students with those of lower ability.

When analyzing bullying status, given that the total number of students suffering from bullying is low, most of the teams formed by teachers consist of two students, neither of whom suffer from bullying. However, we find that students suffering from bullying are more often paired with a student who does not suffer from bullying.²⁰

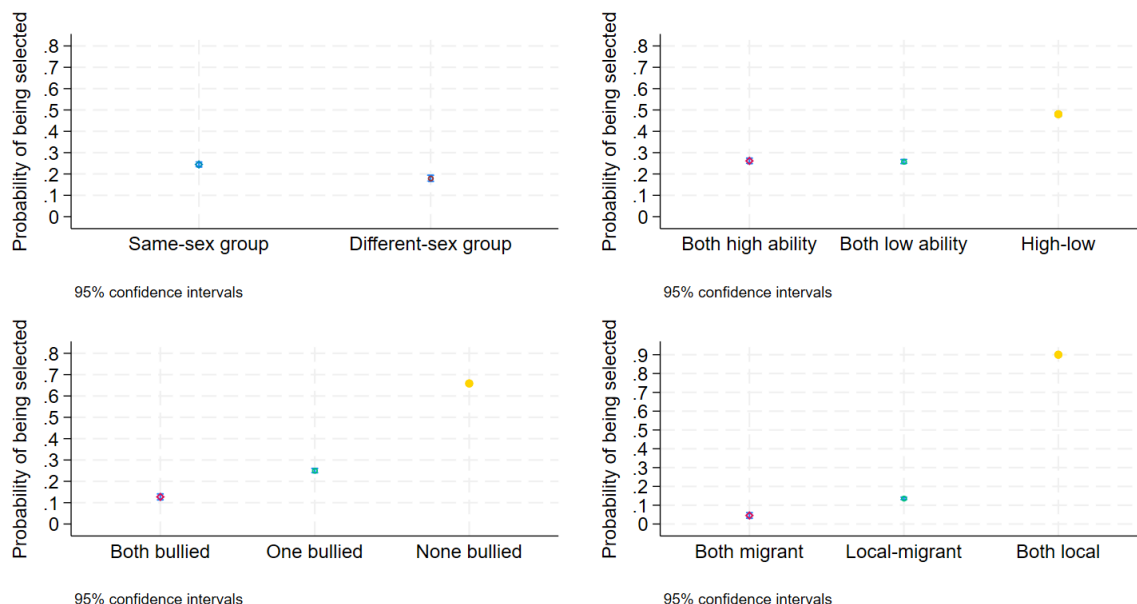


Fig. 2.8.Prob. of being assigned to a group by the teacher depending on characteristics of the student and their partner

Lastly, when analyzing migrant status, we observe patterns similar to those found for bullying. Given the limited number of migrant students in our sample, most teacher-formed teams consist of two local students.²¹ However, when focusing on migrant students specifically, we find that they are predominantly paired with local students, while the number of teams comprising two migrant students is very small. This again contrasts with the scenario where students select their own teammates. In such cases, being a migrant has a similarly negative impact to that of suffering from bullying, which would lead us to expect most migrant students to end up in teams consisting only of migrants. In contrast, teachers tend to pair migrant students with local students, likely as a deliberate effort to promote integration.

In summary, while teachers' group formation does take some sociodemographic factors into account, their approach appears to diverge from students' preferences in certain key aspects, such as the prioritization of mixed-ability groups and the integration of marginalized groups.

The next analysis in this section incorporates two distinct indicators: social preferences and friendship relationships. For the first indicator, we classify teacher-formed groups into three categories: both students in the group are kind, both are unkind, and mixed groups. Students are categorized as kind if they demonstrated altruistic or egalitarian behavior during the dictator game, and as unkind if they exhibited selfish or spiteful behavior. The results, presented in Figure 2.9, reveal that teachers form more groups of two kind students together rather than mixing them with unkind students, although these results appear to

²⁰It is important to remember that the bullying status is self-reported by the students, who can report about themselves and their classmates bullying situation. The analysis of bullying status within teacher-formed groups is based on 1,709 observations, of which 15% (257 students) involve students reported as suffering from bullying.

²¹The analysis of migrant status within teacher-formed groups is based on 987 observations, of which 5% (52 students) are migrant students, while the remaining are local students.

stem primarily from the fact that the sample contains a higher proportion of kind students compared to unkind ones. Likewise, when students selected their own groups, social preferences seemed to have little influence on their decisions, which suggests that social preferences are overlooked in the group formation process, at least to a some extent.

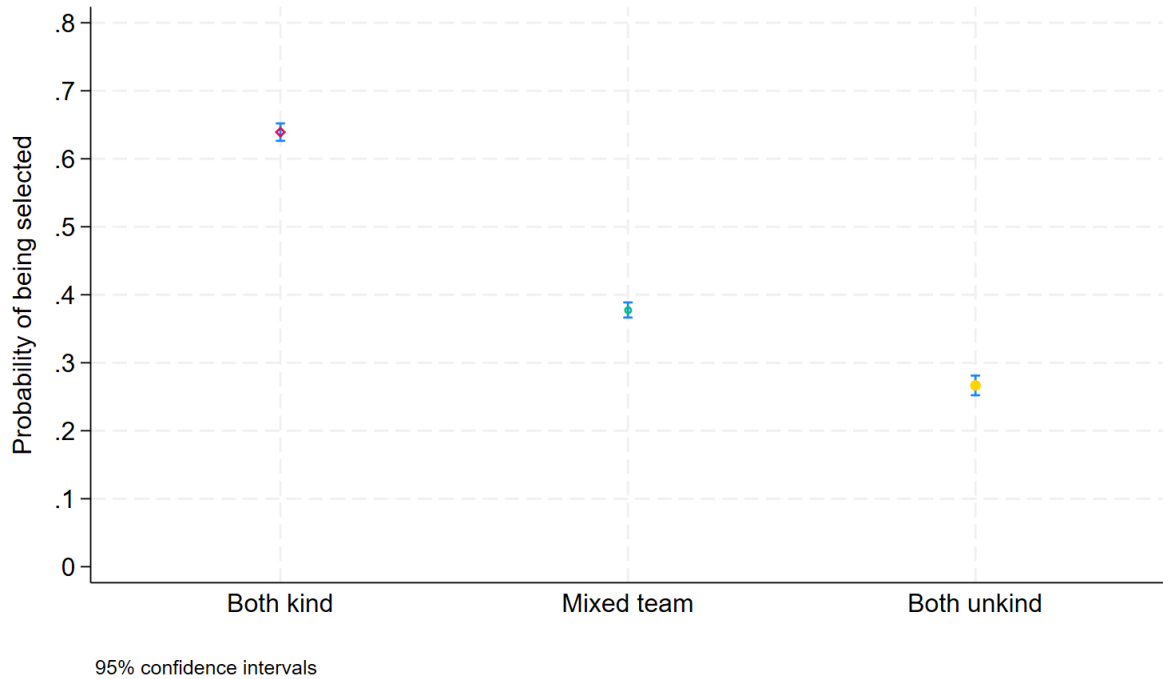


Fig. 2.9. Prob. of being put into a group by the teacher depending on social preferences of the student and their partner

When analyzing the relationship between students assigned to the same group by the teacher, we classify these relationships into three categories: whether the partner assigned by the teacher is someone the student selected as a friend, as an enemy, or neither. The results, shown in [Figure 2.10](#), indicate that teachers predominantly pair students with their friends. This aligns with the results in the random allocation situation, where approximately 40% of the groups formed consisted of a partner the student had selected as a friend.

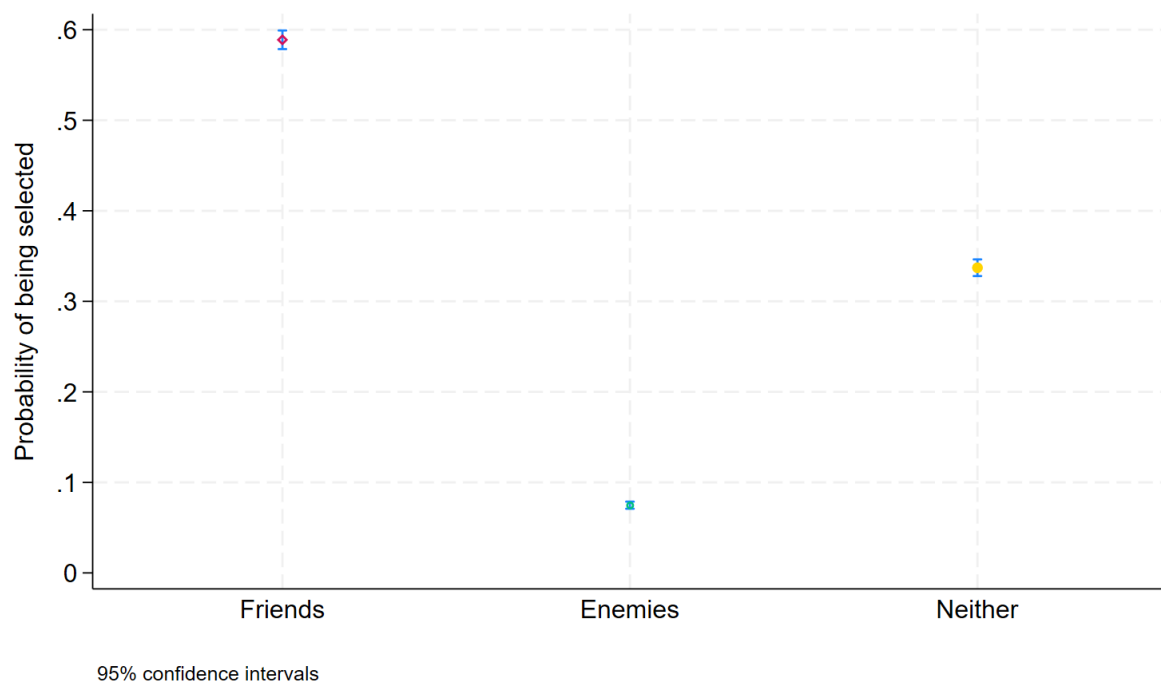


Fig. 2.10. Prob. of being put into a group by the teacher depending on friendship relation between the student and their partner

However, in the random allocation scenario, the number of groups comprising enemies was significantly higher compared to teacher-formed groups. Additionally, the proportion of groups where there was no defined relationship between the students was lower in random allocation than in the teacher-formed groups. These patterns suggest that teachers are aware of the friendship networks within the classroom and take them into account when forming groups, particularly by avoiding pairing together students who do not get along.

In the final analysis of this section, we directly compare the groups formed by teachers with the groups that could have been formed based on the students' stated preferences. Specifically, we categorize teacher-formed groups according to whether the partner assigned to a student was someone the student had selected in the group preferences survey as a preferred group member, someone they would prefer not to work with, or none. These results are presented in [Figure 2.11](#).

Our findings reveal that, in most cases, teacher-formed groups do not align with the students' preferred choices. While there are instances where students are paired with their preferred partners, this is not the norm. However, it is noteworthy that teachers also avoid pairing students with individuals they explicitly indicated they would not want to work with. This suggests that teachers are at least aware of negative relationships among students and deliberately avoid pairing students who would prefer not to collaborate.

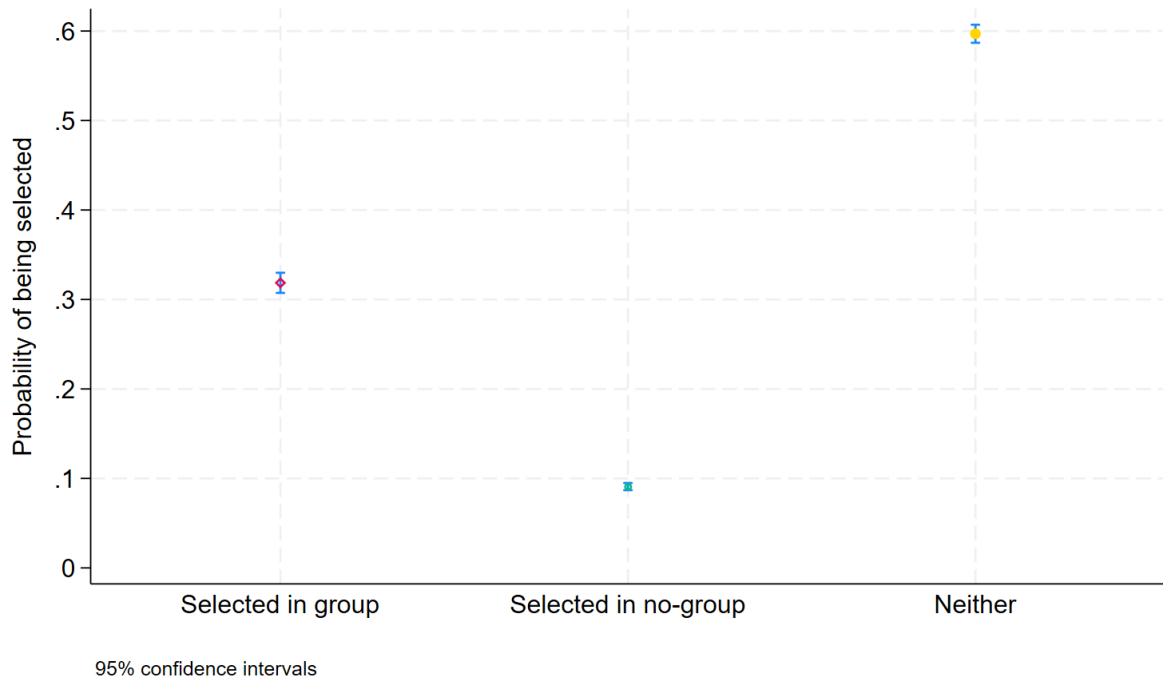


Fig. 2.11. Prob. of being put into a group by the teacher depending on group preferences of the student

The mismatch between teacher-formed groups and student preferences could be attributed to two potential factors. First, teachers may lack detailed knowledge of students' positive preferences, focusing instead on avoiding negative pairings. Alternatively, teachers may intentionally disregard certain student preferences, prioritizing their own criteria for group formation, which could include factors such as academic balance, social integration, or classroom dynamics. This divergence highlights the potential trade-offs between optimizing for student preferences and broader educational objectives when forming groups in the classroom.

2.6 Summary and Concluding Remarks

This paper examines the dynamics of group and friendship formation among students and teachers, focusing on how individual characteristics, such as gender, ability, migrant status, and bullying status, shape the process of selecting partners for collaborative work in schools. Using data from a large-scale field experiment conducted across schools in the Barcelona metropolitan area, in which we collected data about 1,735 students, we compare student-formed groups with teacher-assigned groups to investigate how preferences, relationships, and characteristics influence group composition.

Our results reveal distinct patterns in group formation preferences. Students exhibit a strong tendency to form groups to work in class based on sociodemographic traits, particularly gender and academic ability. High-ability students are more likely to be selected, with a 9.7% increase in likelihood per quartile increase in ability, while migrant students and those experiencing bullying face significant barriers, being 25% and 35% less likely, respectively, to be chosen by their peers. Interestingly, gender has no significant effect on overall team selection likelihood. However, a deeper analysis shows strong gender homophily, with both male and female students preferring same-gender partners. These patterns suggest that students prioritize attributes they perceive as beneficial to group performance or social compatibility.

Finally, perceived ability emerges as a key factor in group selection, even outweighing actual ability. Students who believe they are good at math, even when their test results do not support this belief, are 16% more likely to be selected by their peers. This highlights the importance of perceived competence in shaping desirability as a group partner, suggesting that confidence may be as influential as actual skill in collaborative settings. These findings underscore the need for schools to actively foster students' self-confidence, which would imply that all students, but particularly those with lower measured ability, could benefit from more opportunities for collaboration.

When analyzing friendship formation, we find both similarities and differences compared to team formation. Gender homophily remains a dominant factor, but ability plays a less pronounced role. Additionally, migrant status surpasses bullying as a barrier to forming friendships, likely reflecting the long-term nature of friendships compared to the short-term task-oriented nature of teams. Female students are less likely to be selected as friends, a finding driven by their lower average number of selected friends and a stronger preference for same-gender friendships.

Teachers, on the other hand, adopt a different strategy for group formation. While they also reflect some gender-based patterns, their approach prioritizes academic diversity and social integration. Teachers often pair students with differing ability levels (high-ability students with low-ability peers), a strategy that contrasts with the clustering of high-ability students observed in student-formed groups. Teachers also appear to be mindful of classroom relationships, avoiding pairing students with poor interpersonal dynamics, such as those labeled as "enemies". However, in most cases, teacher-formed groups do not align with students' stated preferences, suggesting a deliberate divergence to prioritize educational or social objectives.

Our study highlights important implications for educational practice. Student-formed groups, while reflecting individual preferences, risk perpetuating exclusion for marginalized students and creating ability-based clusters. Teacher-assigned groups, by prioritizing diversity and addressing interpersonal dynamics, offer a counterbalance but may overlook student preferences. Striking a balance between these approaches could optimize both performance and social cohesion. Additionally, fostering environments that challenge perceptions of ability and encourage inclusion may enhance group dynamics and outcomes. Future research could explore the long-term effects of these dynamics on learning outcomes, social development, and classroom relationships, contributing to a deeper understanding of group formation in educational contexts.

2.7 Appendix

A1 Figures

Selecciona els teus amics/amigues. Recorda que pots seleccionar tots els companys del teu curs.

☐ 2º ESO - B
 ☒ 2º ESO - A
 ☐ 1º ESO - B
 ☐ 1º ESO - A

☐ Elijah Ross - 2º ESO A	☐ William Mitchell - 2º ESO A
☐ Sophia Thompson - 2º ESO A	☐ Elizabeth Reed - 2º ESO A
☐ Noah Bennett - 2º ESO A	☐ Michael Nelson - 2º ESO A
☐ Isabella Allen - 2º ESO A	☐ Grace Mitchell - 2º ESO A
☐ Lucas Baker - 2º ESO A	☐ Daniel Parker - 2º ESO A
☐ Olivia Jenkins - 2º ESO A	☐ Emily Green - 2º ESO A
☐ Liam Wood - 2º ESO A	☐ Samuel Phillips - 2º ESO A
☐ Ava Martinez - 2º ESO A	☐ Scarlett Wilson - 2º ESO A
☐ Benjamin Turner - 2º ESO A	☐ James Carter - 2º ESO A

Fig. A2.1. Example of the screen students were shown when asked to select their friends. They can select people from all classes within the same grade level. Question asked changed, asking to select friends/enemies (people they do not have a good relationship with)/people they want/do not want in their team. Names are fictional.

Selecciona els teus millors amics/amigues

☐ Daniel Parker - 2º ESO A	☐ Henry Powell - 2º ESO B
☐ Lucas Baker - 2º ESO A	
☐ Isabella Allen - 2º ESO A	
☐ Chloe Patterson - 2º ESO B	

Fig. A2.2. Example of the screen shown directly after the previous one, with a subsample of the list showing only those selected in the previous case, and asking now to select from this subsample their best friends/worst enemies (people they like the least), etc. Names are fictional.

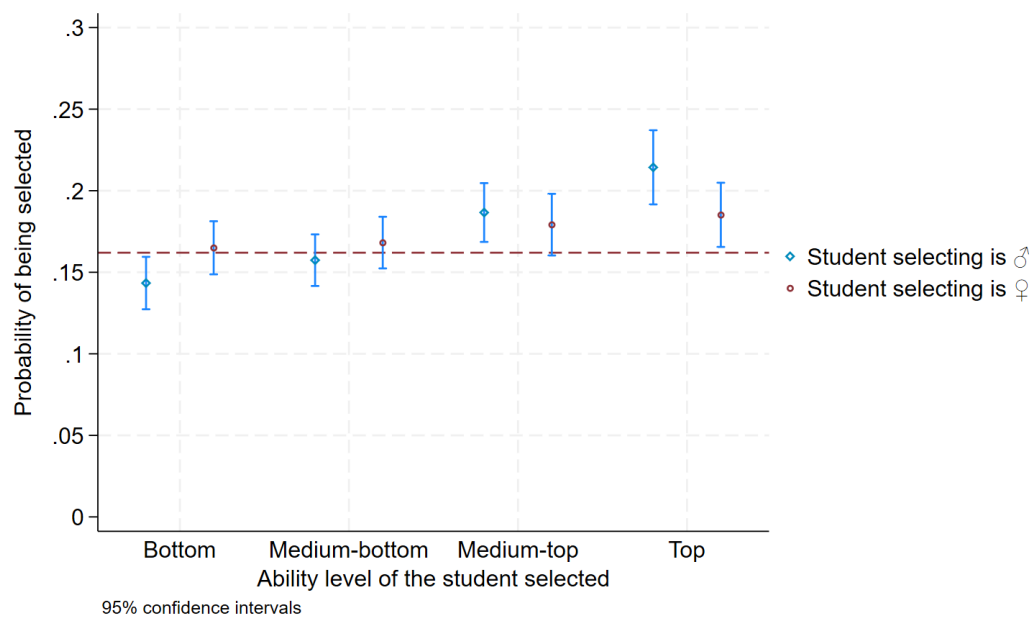


Fig. A2.3. Probability of being selected depending on gender of the student selecting and ability level of the student selected

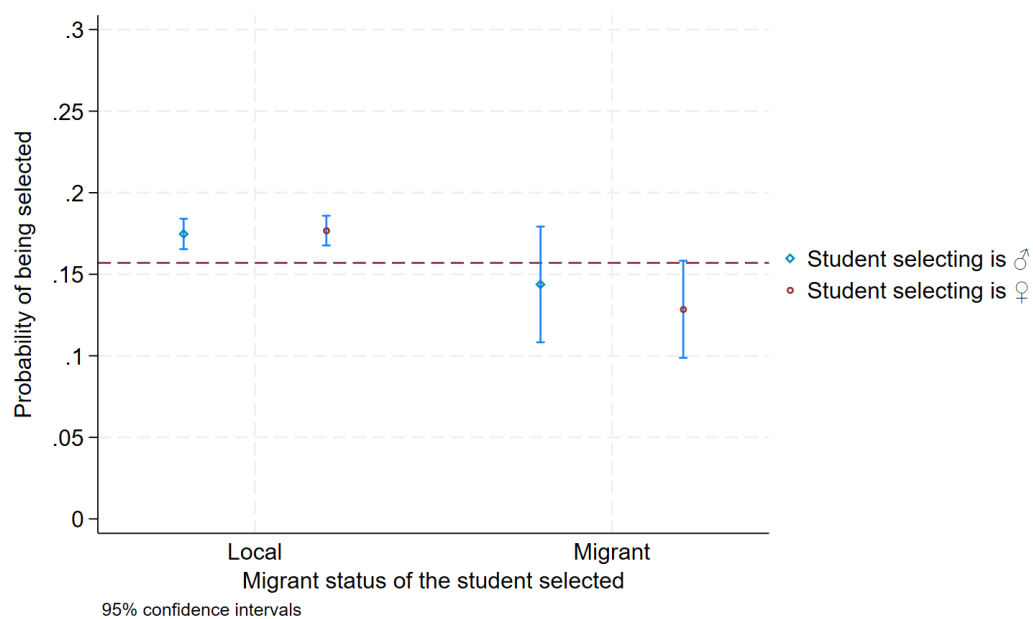


Fig. A2.4. Probability of being selected depending on gender of the student selecting and migrant status of the student selected

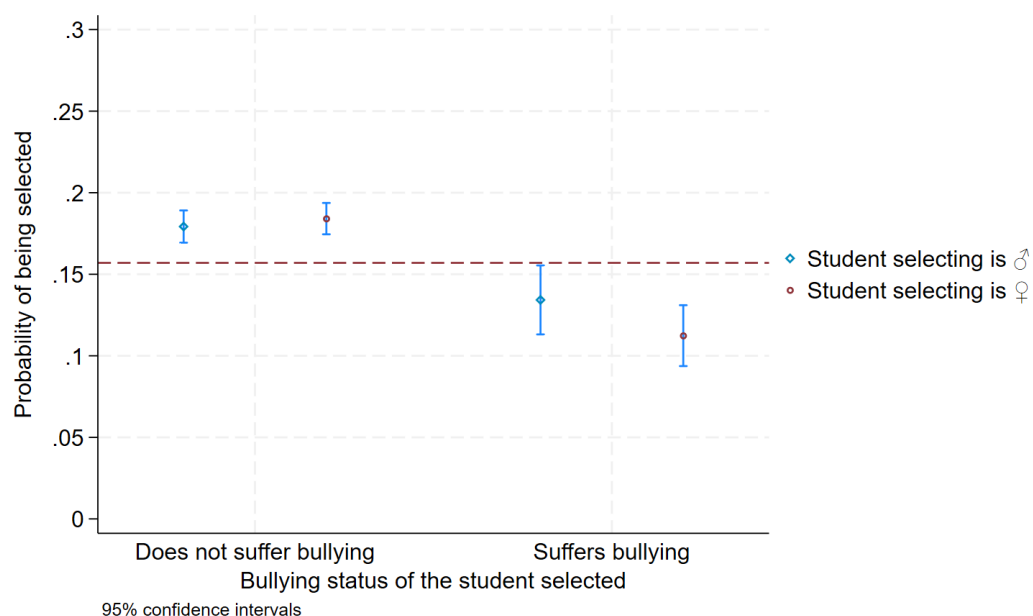


Fig. A2.5. Probability of being selected depending on gender of the student selecting and bullying status of the student selected

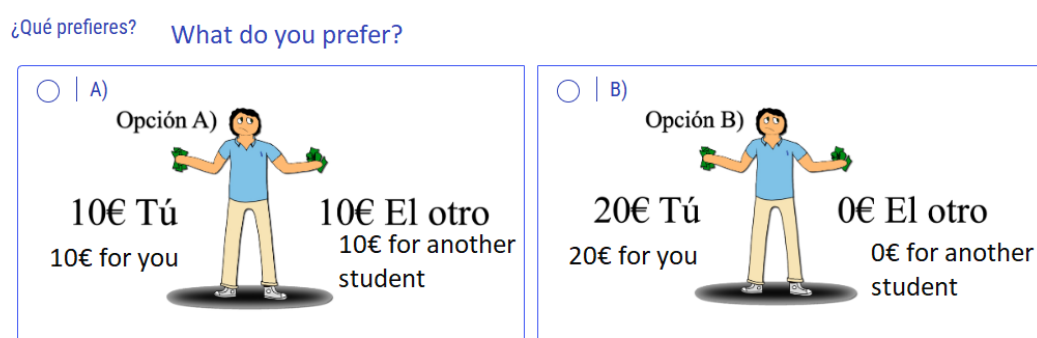


Fig. A2.6. Example of the screen students were shown when asked to play the dictator game, option X. Order was random.

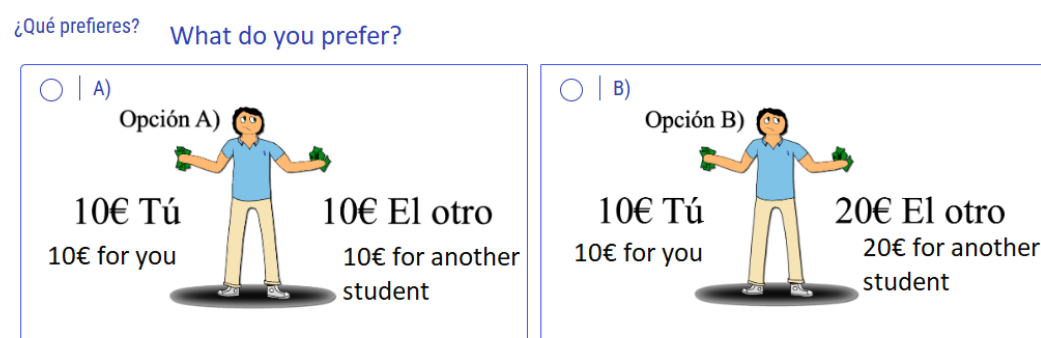


Fig. A2.7. Example of the screen students were shown when asked to play the dictator game, option Y. Order was random.

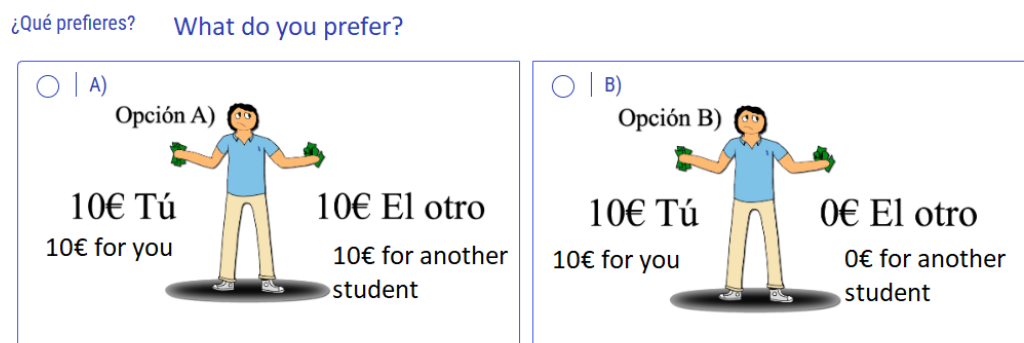


Fig. A2.8. Example of the screen students were shown when asked to play the dictator game, option Z. Order was random.

A2 Tables

Table A3.1: School characteristics.

School	City	Type of school	Socioeconomic index (IST)	Diff. with Catalonia avg. (100)
1	Badalona	Charter	109.9	9.9
2	Badalona	Charter	83.2	-16.8
3	Barcelona	Public	108.4	8.4
4	Barcelona	Charter	102.1	2.1
5	Barcelona	Public	107.2	7.2
6	Barcelona	Charter	121.3	21.3
7	Barcelona	Public	123.6	23.6
8	Barcelona	Public	122.6	22.6
9	Barcelona	Public	109.9	9.9
10	Barcelona	Charter	105.6	5.6
11	Barcelona	Public	101.8	1.8
12	Barcelona	Public	107.2	7.2
13	Ripollet	Public	110.7	10.7
14	Sabadell	Charter	116.8	16.8
15	Sabadell	Public	84	-16.0
16	Sant Quirze del Vallès	Public	124.2	24.2
17	Santa Coloma de Gramanet	Charter	92.6	-7.4
18	Santa Perpètua de Mogoda	Charter	89.1	-10.9
19	Viladecavalls	Public	115.0	15.0
Average			107.1	7.1

Source: Own elaboration using data provided by the [Institut d'Estadística de Catalunya](#)

Table A3.2: Regression of individual characteristics on perceived math ability

	(1)
	Perceived math ability
Ability	0.228*** (0.024)
Female	-0.476*** (0.059)
Suffers from bullying	-0.120 (0.070)
Migrant	0.142 (0.112)
Class FE	✓
Observations	1653
Mean of Dep. Variable	2.304

Standard errors in parentheses, clustered by class
 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3.3: Regression on the probability of being selected depending on social preferences and perceived ability, by gender

	(1)	(2)
Social preferences	0.00483 (0.00603)	
Female	0.0244 (0.0199)	0.0354** (0.0140)
Social preferences $\times$ Female	-0.0117 (0.00906)	
Perceived math ability		0.0226*** (0.00341)
Perceived math ability $\times$ Female		-0.00965* (0.00521)
Class FE	✓	✓
Observations	1131	1653
Mean of Dep. Variable	0.169	0.165

Standard errors in parentheses
 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3.4: Factors affecting perceived math ability

	(1)
Baseline test score	0.169*** (0.0171)
Female	-0.475*** (0.0604)
Suffers from bullying	-0.110 (0.0715)
Migrant	0.136 (0.116)
Class FE	✓
Observations	1,623
Mean of Dep. Variable	2.304

Standard errors in parentheses, clustered by class. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Perceived math ability is positively associated with actual ability, as indicated by the significant and positive coefficient of the baseline test score. However, we also observe a negative association between perceived ability and being female: female students report lower self-assessed math competence compared to their male peers, even after controlling for actual performance. This finding aligns with prior research documenting gender gaps in math self-efficacy and confidence ([Correll \(2001\)](#), [Else-Quest et al. \(2010\)](#), [Zander et al. \(2020\)](#)).

Table A3.5: Determinants of Team Selection Probability

	(1)	(2)
Perceived math ability	0.0106*** (0.00247)	0.0126*** (0.00263)
Friend popularity	0.0122*** (0.000947)	
Female	0.0218*** (0.00657)	0.0138* (0.00711)
Baseline test score	0.0124*** (0.00246)	0.0143*** (0.00267)
Migrant	−0.0177 (0.0117)	−0.0453*** (0.0124)
Suffers from bullying	−0.0301*** (0.00691)	−0.0560*** (0.00729)
Number of friends		0.00191*** (0.000349)
Class FE	✓	✓
Observations	1,653	1,653
Mean of Dep. Variable	0.165	0.165

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3 The Cost of Friendship: Who Is a Better Working Partner?

3.1 Introduction

Effective teamwork is essential in modern economies, where collaborative efforts across various contexts are crucial to operational success, and used in a variety of situations, as noted by [Wetmore and Summers \(2003\)](#). This skill is increasingly regarded as a core competency by many organizations in their hiring processes. As noted by [Weidmann and Deming \(2021\)](#), over 50% of employers identify the ability to work effectively with others as a critical qualification for workplace performance. This emphasis on collaboration extends beyond the workplace and, as outlined by [Hill \(1982\)](#), group interactions are often favored in a variety of contexts. This idea resonates within educational environments as well, where group projects have become an integral component of students' learning experiences throughout their academic journey. Yet, what should be the primary objective of these group projects? Is their main purpose to foster students' collaborative skills, or should they serve as a medium for shared knowledge acquisition? Beyond just academic performance, what value is attributed to the well-being of students as they participate in classroom teamwork? Finally, to what extent does the specific composition of these groups influence the achievement of these various educational goals?

Previous literature found that team composition can significantly impact individual and group performance. For instance, [Weidmann and Deming \(2021\)](#) discovered that certain individuals consistently enhance their team's performance, acting as key players to collective success. In another study, [Bandiera et al. \(2010\)](#) examined the concept of "Social Incentives in the Workplace" to analyze how productivity may vary based on the social connections among coworkers. Beyond individual contributions, the dynamics within a team, particularly the interpersonal relationships among members, play a crucial role. [Shah and Jehn \(1993\)](#) highlight that these relationships, especially friendships, can impact job performance, suggesting that the quality of social bonds within a team affects how effectively individuals work together. This indicates that both the specific individuals in a team and the nature of their relationships are essential factors in achieving optimal outcomes. In a related context, [Thompson et al. \(1998\)](#) delved into the decision-making processes of groups across various contexts, with a particular focus on how these processes are influenced by the interpersonal relationships among group members. Similarly, [Leaper \(1998\)](#) explored the impact of gender dynamics on decision-making processes between friends, examining how responses to suggestions differ based on the gender of the speaker and their partner, while [Hsieh et al. \(2020\)](#) investigated the interaction dynamics of groups formed by strangers in decision-making scenarios.

However, could these interpersonal dynamics extend to the educational setting as well? Considering the students' ongoing maturation and the tailored incentives in this educational context, these dynamics might play out differently. While these dynamics may unfold differently due to the unique context of education, if the impact of team composition in schools mirrors that seen in professional environments, the choice of group members may have deep implications for the students' academic development. To

explore this possibility, it is important to investigate how students' performance and well-being vary based on their group peers. This involves understanding both the general outcome of group composition and how these effects may differ among students with various characteristics and abilities. For instance, could strategically pairing students based on shared or contrasting traits lead to the most effective outcomes?

This paper investigates these questions through a field experiment conducted in diverse primary and secondary schools. By analyzing both students' preferences in choosing partners and the groups formed by teachers, we assess how relationships within teams influence academic performance and comfort.¹ If group composition has a tangible effect on any of these aspects of students' development, we expect to observe variations in individual and group outcomes, contingent on the specific relational dynamics and characteristics among group members.

To the best of our knowledge, this study is the first to investigate the trade-off between well-being and performance in relation to the characteristics of group members in the school context, although the effects of team composition on performance have been previously studied. For example, [Aparicio Fenoll and Zaccagni \(2022\)](#) explored the impact of letting students choose their partners on group gender composition and individual versus group performance, finding there are no differences in performance between male and female-dominated teams when formed endogenously, but female-dominated teams perform lower than other teams when teams are formed exogenously. [Baccara and Yariv \(2013\)](#) created a model to study how endogenously formed teams of peers affect the final distribution of resources towards a common goal, which could also be applied to the school context: how students with different valuation with respect to their performance and different costs of their effort decide to create teams, and the ultimate allocation (performance) of these teams. In an accompanying paper, [Izquierdo-Nofuentes et al. \(2025\)\(in progress\)](#), we examine how students' precedents and characteristics (relationship, ability, similarity, etc.) affect their choice of teammates in class, finding that students prefer to make teams with those similar to them in terms of gender, and those who are better in terms of ability. Likewise, [de Matos Fernandes et al. \(2024\)](#) investigated the decision-making process students employ in forming groups, focusing on whether these decisions are influenced by similarities in characteristics or by existing close relationships, whereas [Wang et al. \(2017\)](#) explored how the decision-making mechanisms in risky situations vary between groups and individuals, as group decisions impact not only the individual but also their teammates, a factor that may similarly influence how much effort students choose to put into a collaborative task school setting.

Furthermore, little is known about the specific effects of friendship on performance within the school context and its influence on the comfort of students, which could also impact individuals' performance and health. Considering the unique characteristics of the school setting, it is crucial to analyze teamwork with a focus not only on performance but also on student well-being. This paper seeks to address this gap, offering new insights into the factors shaping these dynamics and their varying effects on different types of students. Specifically, we aim to explore the relationship between friendship, ability, performance and well-being. We will examine the potential effects across various student demographics, such as gender (male versus female), ability levels (high versus low), and age, with a particular focus on how these factors influence learning peer effects. If team composition is a factor in determining peer learning, we anticipate observing improvements in individual performance after team interactions for some students, contingent on the characteristics of these teams.

¹In this paper we will refer interchangeably to the groups formed by two students as "groups," or "teams", acknowledging that each term reflects the same composition of only two participants.

In our study, we conducted an experiment across various schools in the metropolitan area of Barcelona, involving a total of 1,735 students. Participants engaged in a two-hour session that included a basic information survey, the construction of a student relationship network, and several assessments. Our survey also incorporated questions related to students' suffering from bullying in class, since this can be related to both performance and well-being. At the same time, teachers were asked to complete a short survey detailing how they would pair students in their class for a given assignment.

Our sample primarily consists of secondary education students aged 12 to 16, alongside students in the final two years of elementary school (ages 10 to 12) and students in vocational training education (aged 16 and above). These students are drawn from both public and charter schools located in various cities and neighborhoods with distinct socioeconomic characteristics.

Our findings indicate that the nature of students' relationships can influence their performance when working together. When only one student perceives a relationship—whether positive (friendship) or negative (enmity)—there is no statistically significant effect on group performance. However, when the relationship is mutual (i.e., both students identify each other as friends or enemies), we find weak evidence (statistically significant at the 10% level) suggesting a decline in group performance. Specifically, when both students selected each other as friends, group performance decreases by approximately 5.7%. This reduction is even more pronounced when both students identify each other as enemies, with performance declining by around 11%. These results suggest that strong interpersonal ties—whether positive or negative—may hinder collaboration, possibly due to distractions among friends or conflict between enemies. This contrasts with results from previous studies in laboratory settings, such as those by [Jehn and Shah \(1997\)](#), who found that groups of friends outperform those comprised of acquaintances on various tasks. Furthermore, as demonstrated in our companion paper [Izquierdo-Nofuentes et al. \(2025\) in progress](#), friendship relations play a critical role in students' choices regarding team formation.

Two additional factors that appear to shape group performance are the individual's own ability and the ability of their assigned partner, with both factors being nearly equally important. Notably, the effect size of these ability-related factors is comparable to the negative impact observed when both students in a pair identify each other as friends rather than having no prior relationship. Compared to their performance when working alone, we observe that, on average, group performance improves. Furthermore, after participating in group work, students' individual performance remains higher than their initial levels, even though it is lower than their performance in the group setting, suggesting students are able to learn from each other when working together. A closer examination reveals that students with a score below the median in the baseline test (referred to as "low-performing" students) benefit significantly from group work; their performance is enhanced when paired with another low-performing student and improved even further when paired with a student whose score in the baseline test is above the median (referred to as "high-performing" students). In contrast, high-performing students experience a decrease in performance when working with a low-performing student. When paired with another high-performing student, their performance is better than when paired with a low-performing student, but it continues to be lower than their initial levels. In summary, low-performing students gain from group work regardless of their partner's ability, while high-performing students encounter diminished performance in group settings irrespective of their partner's ability. Compared with the results found in [Saleh et al. \(2005\)](#), some similarities can be found. In their study, low-ability students perform better when grouped with students with higher ability than them. In contrast, our findings indicate that low-ability students exhibit improved performance regardless of their partner's ability. Additionally, [Goldman \(1965\)](#) reported similar outcomes, highlighting that group performance consistently surpassed individual performance. Specifically, he found that low-performing students enhance their performance when paired with a high-performing student and

The Cost of Friendship: Who Is a Better Working Partner?

that neither high-performing nor low-performing students show performance changes when working with a student of the same ability level. However, our findings diverge as we found that low-performing students also improve their performance when working with another low-performing student, and high-performing students exhibit decreased performance irrespective of their partner's ability, with a more pronounced decline when paired with a low-performing student.

Upon further examination of performance effects, gender emerges as another influential factor, albeit much less so than ability. Female students experience greater benefits from group work compared to male students, showing a larger improvement in their group performance relative to individual performance. Furthermore, students who experience bullying tend to perform worse on average, both individually and in groups, with performance unaffected by partner characteristics. Meanwhile, wealth and country of origin show no significant impact: performance remains consistent regardless of the student's or their partner's wealth level, or whether they were born in Spain or abroad.

In addition to performance, it is important to consider the well-being effects of group composition in a school setting, where, in our study, well-being is specifically measured by how comfortable students feel when working in groups, depending on their partners. When students decide who to group with, their choice may be driven by a combination of the expected well-being and expected performance they anticipate from working with a specific partner, with varying weights depending on individual student characteristics. This emphasis on well-being could explain why friendship plays a significant role in students' group selection, even though it negatively impacts team performance. Our analysis shows that, on average, girls report greater comfort in group settings than boys, particularly when paired with other girls. Conversely, students who experience bullying are generally less comfortable in groups, especially when paired with another student who also suffers from bullying. Our investigation into the impact of relationships highlights a positive effect of having a good relationship with group partners on students' well-being. Notably, students working with individuals they do not favor exhibit significantly lower well-being levels compared to other scenarios. Additionally, albeit to a lesser extent, students working with friends report significantly higher well-being than those working with neutral acquaintances. However, as observed in the analysis of performance, this increased well-being does not necessarily lead to improved performance. Our findings suggest that while students feel more comfortable working with friends, these pairings result in lower group performance. This trade-off between well-being and performance highlights a fundamental tension in group dynamics. While friendships create a sense of ease and emotional security, they may also reduce efficiency. Friends might hesitate to challenge each other's ideas, engage in more off-topic discussions, or feel less pressure to perform at their best. As a result, while these groups report higher comfort levels, their performance tends to suffer in comparison to groups composed of neutral acquaintances. This finding suggests that students may not always prioritize academic success when selecting partners, instead valuing the emotional benefits of working with familiar peers. Moreover, we observed a positive and statistically significant relationship between students' reported level of comfort and their subsequent responses to morale state questions. These responses, gathered at the end of the session, suggest that the comfort experienced during collaborative assignments may influence their perceived morale state. While working with friends may reduce efficiency, it could foster a more positive perception of teamwork, potentially influencing long-term attitudes toward collaboration and learning.

The remainder of the paper is organized as follows. [Section 2](#) provides background information about the experiment. [Section 3](#) describes the experimental design and implementation. [Section 4](#) presents the data collected. In [Section 5](#), we present the results of our analysis. Finally, [Section 6](#) summarizes and concludes.

3.2 Background

Prior to conducting the experiment, we preregistered our primary hypotheses and methodological approach with the [Registry of Efficacy and Effectiveness Studies](#). Additionally, conducting research with minors requires adherence to strict protections and laws, which is why we sought and received external ethical approval from the [Bioethics Commission](#) at the University of Barcelona,² which formally approved the study on November 2nd, 2022. Furthermore, all responses were guaranteed to be anonymized, ensuring that researchers are unable to identify the participants. To address challenges associated with non-standard samples and to reduce the incidence of missing data, we streamlined the format of responses by primarily employing multiple-choice questions in lieu of open-ended ones. The software's design made it compulsory for participants to respond to all questions before proceeding to the next one, thereby preventing the omission of answers. Nevertheless, for subjects that might be perceived as sensitive, participants were offered the option to select "I would prefer not to answer."

We conducted experimental sessions in 19 schools located in the greater Barcelona area. Initially, we reached out via email to all schools receiving funding from the Education Department to identify those potentially interested in participating. After explaining the objectives of our research and meeting with several school representatives, we finalized the list of schools that would comprise our sample. The participant pool was recruited through agreements with school headmasters, who consented to conduct the experimental session as a classroom activity integrated into their pedagogical curriculum, which resulted in a high level of participation. Prior to the experimental sessions, the schools provided us with signed consent forms, agreeing to conduct the session as an official school activity, along with class lists detailing the enrolled students. Using these lists, we created unique usernames and passwords for each student in the software employed for the experiment. We then randomly paired students using randomization software, forming the groups that would collaborate during the session. For classes with an odd number of students, the remaining student was designated as a "wildcard" to be assigned on the session day if any pre-assigned partner was absent. In cases where students ultimately worked alone, they were excluded from the group performance analysis.³

The experimental sessions were conducted on-site at the schools, using an online software named Social Analysis and Network Data ([SAND](#)),⁴ which assisted in mitigating privacy concerns. Students could use any of their devices (laptop, classroom computer, tablet or smartphone) in order to access and carry out the experiment. The questionnaire was administered entirely in Catalan. Due to restrictions on experiments involving real money that are in place in schools, good performances or effort were not rewarded. However, students were informed that test results would be sent to their teachers and they could decide how to use them.

²Institutional Review Board IRB00003099.

³The number of students who ended up working alone was very small (36 students, 2% of our total sample), which informed our decision to exclude them rather than using them as a control group for comparison against those who worked in pairs.

⁴This software was commissioned by an amalgamation of public universities and institutions with the purpose of obtaining data from teenagers in schools across Spain. For more information we refer the reader to [Vasco et al. \(2025\)](#).

3.3 Experimental design

The experiment is structured into three distinct phases, as outlined in [Fig. 3.1](#).⁵ During the first phase, students complete a survey designed to gather comprehensive information on their sociodemographics, family background, self-assessed mathematical competence, and academic performance in the previous years. Subsequently, participants are prompted to indicate the level of friendship⁶ they share with their classmates and to select individuals they would prefer to collaborate with on a school task. Screenshots of the questions pertaining to students' social networks are provided in [Fig. A3.1](#) and [Fig. A3.2](#) in the Appendix. This information enables us to analyze decision-making processes regarding friendship and partnership selection, the characteristics prioritized by students in forming these relationships, and the possible differences when selecting a friend instead of your preferences about your group partners.⁷ At the end of the first phase, students are required to complete a series of timed multiple-choice tests in Mathematics⁸ individually, which will be used as a baseline to compare with the subsequent tests. Screenshots of some of these questions are detailed in the Appendix, within [Fig. A3.3](#) through [Fig. A3.19](#).

In selecting questions for evaluating performance, it was essential to ensure they met three key criteria. Firstly, the tasks had to be comparable to activities students typically encounter in school, ensuring relevance and accessibility. Secondly, it was crucial that the tasks could be administered to both individuals and groups, allowing for the assessment of group performance while accounting for individual abilities. Lastly, the questions and their corresponding answers needed to be impartial, allowing us to objectively evaluate both individual and group performance accurately. The tasks used in this experiment meet these criteria. For the mathematics tasks, students are presented with a mathematical problem that may involve a geometric figure, a mathematical operation, or an arithmetic question, among other types. Each problem is accompanied by four potential answers.

The second phase commences with the random assignment of each student to a peer,⁹ with the paired students tasked to jointly undertake another set of multiple-choice tests, in which they are allowed to discuss how to solve the problem to come up with the potential correct solution. Upon completion, participants are separated and asked regarding their comfort levels when working with their specifically assigned partner.

In the final phase, students individually complete a third battery of multiple-choice tests. Additionally, they are requested to answer some questions about their morale state and to identify any classmates, including themselves, who may be experiencing bullying. With this information, we are able to explore the relationship between performance in the different tests and the individual characteristics of the student and their partner.

⁵The analysis of phases 2 and 3 of the experimental session is addressed in the following chapter.

⁶Participants were first instructed to identify individuals in their class whom they regard as friends. From this group, they were then asked to select those they consider their best friends. Subsequently, participants were instructed to indicate individuals with whom they do not maintain a positive relationship (for the purposes of this paper, these individuals are referred to as “enemies” to facilitate reading). From this list, they were then asked to select those with whom they have the most difficult relationships.

⁷For a detailed analysis of these results, readers are referred to [Izquierdo-Nofuentes et al. \(2025\)](#) ([in progress](#)).

⁸Mathematics tests questions were classified according to the grade of the student. For elementary school students, we use questions extracted from [Proves Cangur](#) corresponding to the US-equivalent 5th and 6th grades. For the remaining students, we use questions based from the [ASVAB](#) test, and classify them by grade according to school syllabi.

⁹We run a robustness check to ensure the randomization of the students into groups was properly done, and that the characteristics of students working together are not related.

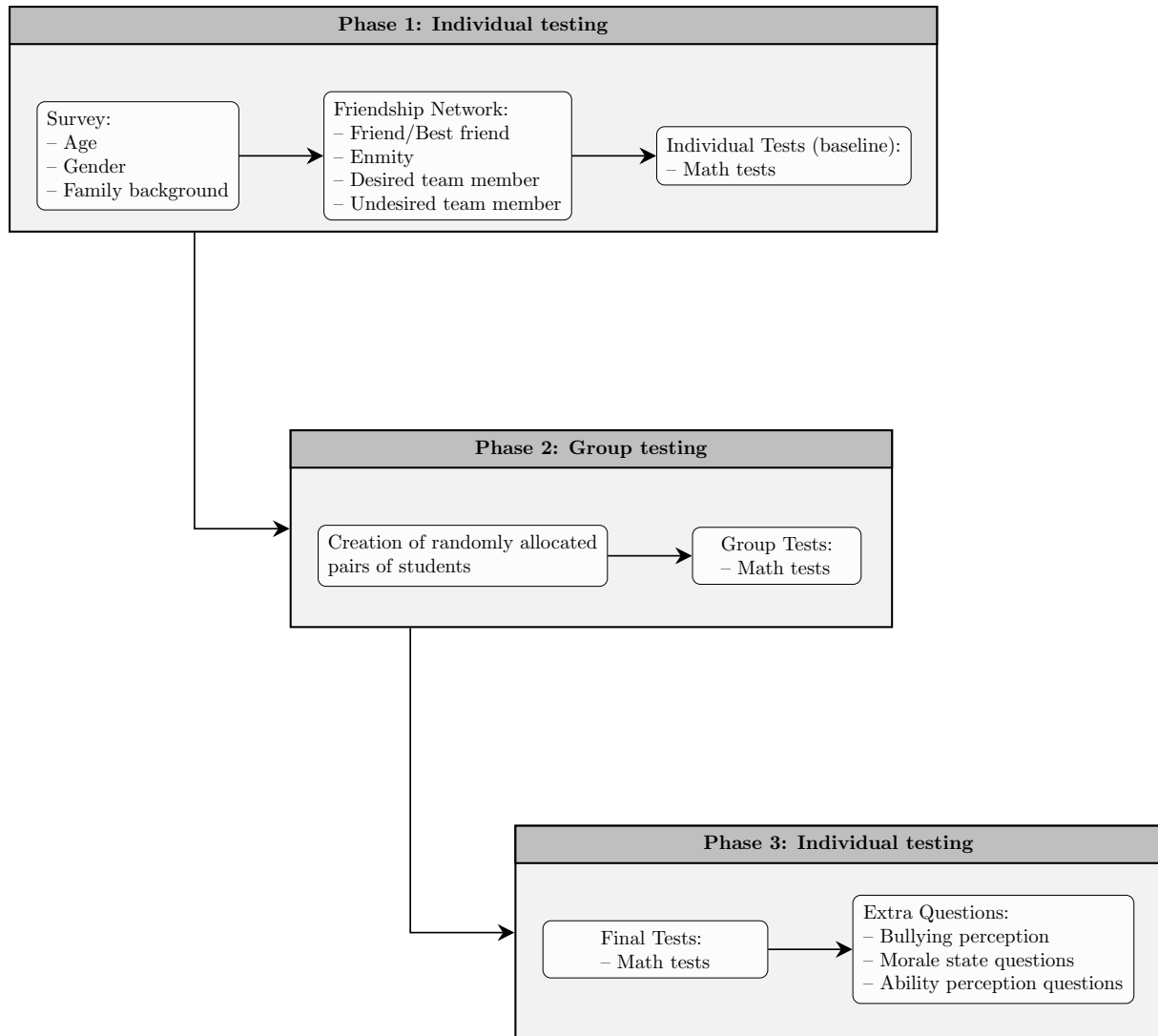


Fig. 3.1. Chronology of the experimental sessions.

Simultaneously, data regarding teachers' preferences for group formation is collected. Specifically, teachers are asked to create hypothetical groups for a classroom task, as well as the reason behind forming these specific groups.

During the experimental session, students are organized by grade and class and seated according to their usual classroom arrangement. A set of instructions, presented in Catalan, is displayed on their screens and read aloud to the students: "Dear students, we are carrying out a lab-in-the-field experiment in the field to understand how social relationships affect group performance. Please, work in silence and answer the questions truthfully." Subsequently, students receive guidance on how to access the experimental platform and the procedure for identifying their randomly assigned partner in the group phase.

Students do not receive feedback on their performance at any point during the session. The tasks are computer-based, with each student utilizing a digital device (be it a computer, laptop, smartphone, or tablet) to complete the tasks. Each participant is assigned a unique username and password, enabling them to respond independently on their device, particularly if they disagree with their partner during the group

phase.¹⁰ Additionally, we record the time at which each student answers each question. This timing data allows us to monitor instances of students potentially selecting random answers and to identify whether participants are collaborating simultaneously with their assigned partners.

3.4 Data and descriptive statistics

Our sample is composed of students from nineteen educational centers, including elementary schools, secondary schools, and vocational training schools, located in the vicinity of Barcelona, with a total of 1,945 potential participants. All public and charter schools in the area were contacted to offer the possibility of participating. The final sample consists of 86 classroom groups and 1,735 students, following the exclusion of those absent on the day of the experiment or who were unable to be paired.

Each experimental session lasted for up to two consecutive hours, although students were permitted to finish earlier. No breaks were allowed during the sessions, and responses were recorded up until the point where each student completed the last question, if time constraints prevented them from finishing all of them. Most participants completed at least the first two tests, and many also completed the final test. This information is synthesized in [Table 3.1](#) below.¹¹

Table 3.1: Student distribution across tests by gender, ability, and bullying status

Test Phase	Male	Female	High-performing	Low-performing	Bullying	No Bullying	Total
Individual Test	848	851	728	971	254	1445	1699
Group Test	847	850	726	971	253	1444	1697
Final Test	825	831	707	949	249	1407	1656

[Table 2](#) presents the main summary statistics for the primary sociodemographic characteristics of our sample. As previously detailed, the sample consists of students from various educational levels: elementary school students comprise 27% of the sample, secondary school students comprise 53%, and vocational school students comprise 20%. The gender distribution is balanced, with approximately equal numbers of male and female students, and the average age is 14.65 years. Additionally, 15% of the students have been identified by their classmates as experiencing bullying¹², and roughly half of the groups are composed of same-sex groups.

¹⁰A total of 543 individuals, accounting for 32.4% of our sample, answered at least one question differently from their partner during the group phase. This number significantly decreases thereafter, with no pairs differing in their answers for more than four questions.

¹¹We conducted robustness checks to account for potential correlations between student characteristics and attrition during the tests.

¹²[Table A2.23](#) in the appendix contains summary statistics for the main characteristics about the individuals suffering from bullying

Table 3.2: Sociodemographics

	Mean	Median	Std. Dev.	Min.	Max.	N
Age	14.65	14.00	3.26	11.00	46.00	1735
Female	0.51	1.00	0.50	0.00	1.00	1735
Same-sex group	0.51	1.00	0.50	0.00	1.00	1735
Suffers from bullying	0.15	0.00	0.36	0.00	1.00	1735
Migrant	0.07	0.00	0.25	0.00	1.00	1676
Elementary school (10-12 years old)	0.27	0.00	0.44	0.00	1.00	1735
Secondary school (12-16 years old)	0.53	1.00	0.50	0.00	1.00	1735
Vocational school (16-46 years old)	0.20	0.00	0.40	0.00	1.00	1735

Table 3.3 provides descriptive statistics for the math test outcomes. The standardized math score, ranging from 0 to 10, is calculated by adding the number of correct responses, dividing this sum by the total number of questions, and then multiplying by 10. The average score for the math component is 3.23, suggesting that either the overall math proficiency among students is quite low, or the test's difficulty exceeded our initial expectations. Nevertheless, a significant divergence in average scores is observed across the three test phases, with notable improvement during group work, followed by a slight decline when students returned to individual work. Despite this decline, scores remained significantly higher compared to the baseline test.

Table 3.3: Math tests outcomes (0-10)

	Mean	Std. Dev.	Min	Max	Q1	Median (Q2)	Q3	N
Baseline Math	3.23	1.60	0.00	10.00	2.14	3.08	4.29	1699
Group Math	3.62	1.44	0.00	8.57	2.86	3.57	4.29	1697
Final Math	3.74	2.05	0.00	10.00	2.50	3.75	5.00	1656

The characteristics of the schools in our sample are detailed in Table A2.1 in the Appendix. The sample includes 8 charter schools and 11 public schools. Ten schools are situated within Barcelona city, while the remaining nine are located in neighboring municipalities. Of these, four schools are located in areas with a socioeconomic index below the Catalonia average, eight have an index above but close to the average, and the remaining seven have an index well above the average. On average, the schools in our dataset have a socioeconomic index that is 7% higher than the Catalonia average.

3.5 Results

A1 Group composition effects on performance

Baseline results

We begin by analyzing how the individual characteristics of students within the same team, as well as their interpersonal relationship¹³, can influence their performance. To explore this, we perform an Ordinary Least Squares (OLS) regression analysis of group performance based on these group characteristics, employing the following equation:

$$\text{Group Score}_i = \beta_0 + \beta_1 G_{ij} + \beta_2 X_i + \beta_3 X_j + \alpha_g + \varepsilon_i \quad (3.1)$$

where the dependent variable is the score obtained by the student during the group phase. G_{ij} represents the characteristics of the group (friendship relation, same-sex group) and X_i and X_j denote the individual characteristics of the student and their partner (baseline test score, gender, wealth, bullying status). The results from estimating equation 1 are presented in Table 3.4. Throughout columns 1 to 5, we gradually alternate between the different group and individual characteristics analyzed, and add additional controls. Specifically, column 1 presents a regression of the friendship relationship among group members on their scores. Column 2 shifts focus to the effect of the group's gender composition (same-sex versus different-sex groups). In column 3, we regress baseline test scores against group test scores. Column 4 combines the previous three models, and in column 5 we add additional controls with individual characteristics of the students (gender, wealth level and bullying status). We report complete results for the last model in Table A2.2 in the Appendix.¹⁴

When students work together on the group phase test, the analysis indicates a negative, albeit only statistically significant at the 10% level, effect between friendship/enemy relation and group performance, but only when the relationship is mutual (i.e. both students selected each other as friends or enemies). The main statistically significant effect identified is related to previous scores on the baseline test, which reflect the initial ability of both the student and their partner.¹⁵ Individual or partner personal characteristics do not exert a direct, statistically significant impact on group performance, and the influence of student ability remains consistent across all model specifications. Moreover, the effect is approximately equivalent for both the individual's own ability and that of their partner. Specifically, an increase of one question in a student's baseline test score is associated with a 5.41% enhancement in their group score, corresponding to a 0.136 standard deviation rise. Similarly, a one-question increase in the partner's score leads to a 4.75% increase in the group's score, translating to a 0.120 standard deviation improvement. Therefore, contrary to findings in previous literature such as Shah and Jehn (1993), which suggest a positive relationship

¹³In this paper, "relationship" refers to the degree of friendship categorized by the students: friends, best friends, enemies, worst enemies, and no relationship.

¹⁴We extended our analysis by including the number of friends a student selected and the number of times they were selected as a friend as explanatory variables. These metrics serve as proxies for the student's social abilities, an important factor to explain performance as shown in Weidmann and Deming (2021). However, in contrast to their findings, our results reveal no significant effect of either variable across any of the model specifications, while the primary results of our analysis regarding the importance of initial ability and mutual friendship/enmity remain robust and unaffected by the inclusion of these additional variables. Detailed results are provided in Table A2.25 in the appendix.

¹⁵The variables "Score baseline test" and "Score baseline test partner" present a correlation of 11.88%.

between friendship and performance, our study reveals a negative association between interpersonal relationships within groups and group performance. Specifically, when two friends work together, the group's performance decreases by 5.68%, equivalent to a 0.323 standard deviation reduction. Similarly, when two students with a poor relationship work together, the group's performance drops by 11%, corresponding to a 2.48 standard deviation decrease.

Table 3.4: Regression of individual characteristics and relationship types on group performance

	(1)	(2)	(3)
Friendship both ways	-0.147* (0.110)	-0.202*** (0.103)	-0.207* (0.109)
Friendship one-way	-0.139 (0.109)	-0.150 (0.104)	-0.150 (0.109)
Enmity with friendship	0.0119 (0.323)	0.123 (0.320)	0.114 (0.317)
Enmity one-way	-0.0388 (0.151)	-0.0737 (0.156)	-0.0714 (0.157)
Enmity both ways	-0.490** (0.230)	-0.418* (0.219)	-0.401* (0.215)
Score baseline test		0.199*** (0.0254)	0.196*** (0.0264)
Score baseline test partner		0.177*** (0.0270)	0.173*** (0.0274)
Class FE	✓	✓	✓
Controls	×	×	✓
Observations	1635	1635	1635
Mean of Dep. Variable	3.645	3.645	3.645

Standard errors in parentheses, clustered by class. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Omitted category is "No relationship between students"

When comparing these effects to ability-related factors, the negative impact of mutual friendship within a pair is approximately equivalent to a one-point reduction in a student's baseline test score (which we use to measure their initial ability). Meanwhile, the effect of mutual enmity is even more pronounced, roughly equivalent to a two-point reduction in baseline test performance. This comparison highlights that relationship dynamics within a group can significantly influence performance, at times exerting an impact comparable to differences in individual ability. When compared with the effects of ability, working with a partner when both students consider each other as friends is roughly equivalent to a reduction in one point in the score of the baseline test (a reduction in the initial ability of the student), and working in a group

The Cost of Friendship: Who Is a Better Working Partner?

in which both students consider each other as enemies is equivalent to a reduction in two points in the score in the baseline test. In cases where the relationship is one-sided, a negative effect is still observed, although it is not statistically significant. These results provide weak evidence that working with either friends or individuals with whom one has a strained relationship is detrimental to group performance. The negative effects can be attributed to several factors: individuals who dislike each other may avoid communication and collaboration, leading to reduced performance—an effect roughly twice as large as the impact of their individual ability. Similarly, friends may spend excessive time discussing personal matters rather than focusing on the task, resulting in a performance decline comparable to the effect of their ability. Thus, in terms of performance outcomes, the optimal group composition appears to be one in which members do not have a prior personal relationships.

In the next analysis, we investigate how baseline scores are influenced by individual characteristics and how final phase test scores may be affected by both individual characteristics and those of the partner during the group phase. The results corresponding to the baseline test are presented in [Table 3.5](#). Our findings indicate that while a student's wealth level and country of origin¹⁶ does not significantly impact their scores, gender and bullying status do exert a significant influence.

Specifically, female students and those experiencing bullying score, on average, 0.363 points lower on the individual math test compared to their counterparts, which translates into approximately 0.23 standard deviations below the mean. This gap represents about an 11% decrease from the average score of 3.24, underscoring a significant disparity in performance. For the final test scores, results are detailed in [Table A2.3](#). Consistent with the group score analyses, the only statistically significant effects derive from the students' own abilities. The partner's ability during the group phase has a positive effect on the final test score, although this effect fails to reach any usual level of statistical significance. Female students also present a lower score on the final test than male students, although this difference is only statistically significant at the 10% level.¹⁷ Additionally, migrant students score lower on the final test than local students, with this difference being statistically significant at the 5% level.

¹⁶In our case this means either Spain or abroad.

¹⁷Part of the observed gender effect on both the group and final phase tests is accounted for by the effect of ability, as measured by the baseline test score. Female students perform worse on this test, and this score has a significant impact on the performance on both the group and final tests.

Table 3.5: Regression of individual characteristics on baseline performance

	(1) Score baseline test
Female	-0.365*** (0.0848)
Wealth	0.0420 (0.0683)
Suffers bullying	-0.340*** (0.114)
Migrant	-0.181 (0.153)
Class FE	✓
Observations	1627
Mean of Dep. Variable	3.246

Standard errors in parentheses, clustered by class.
 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Heterogeneous analysis

To gain deeper insights into how various student characteristics influence test scores, we conduct five separate heterogeneity analyses, each focusing on a distinct characteristic of those collected in our dataset: ability, friendship, gender, bullying status and migration status.¹⁸

In our baseline investigation of the factors potentially affecting group performance, initial ability —defined by the baseline test score— was the factor that exhibited the strongest statistically significant effect. To analyze this relationship more closely, we categorize students based on whether their baseline test score is above or below the sample median, labeling them as 'high-performing' or 'low-performing' students.¹⁹

Afterwards, we calculated the average score across the three test phases, segmenting by both the student's type and their partner's type.²⁰ The results are presented in [Figure 3.2](#).

¹⁸We also conduct an analysis by wealth level. Due to the complexity introduced by multiple categories, we omit visual representation but include the detailed ANOVA and Tukey's Honestly Significant Difference (HSD) test results in the Appendix, in [Table A2.19](#) and [Table A2.20](#). These results indicate no significant differences across any wealth groups.

¹⁹As previously shown, the sample median score is 3.08. Students scoring above this median are considered "high-performing," and those scoring at or below it are considered "low-performing."

²⁰The baseline score is considered independently of the partner's type, since the test is carried out individually. We examine differences in individual scores depending on the partner type in the subsequent group phase, finding no significant differences, as expected. For the final phase, we account for both the student's and partner's type to assess any residual impact from prior collaboration on subsequent individual performance.

In the group phase, our analysis reveals distinct performance patterns based on student composition. Groups consisting of two high-performing students achieve higher performance levels than those composed of two low-performing students, as could be expected. Mixed groups, where one high-performing student is paired with a low-performing student, displayed performance outcomes that lie between these two extremes.²¹ Interestingly, high-performing students exhibit a decline in performance during the group phase compared to their individual phase outcomes, irrespective of being paired with a fellow high-performing student. Conversely, low-performing students demonstrate an improvement in their outcomes during the group phase, even when partnered with another low-performing student. This suggests that low-performing students reap substantial benefits from collaborative work.²²

Examining the final test results provides additional insights. Both high-performing and low-performing students maintain performance levels similar to those observed during the group phase. Notably, when compared to the initial test, low-performing students appear to sustain the improvements achieved through group work, performing above their baseline test levels. In contrast, high-performing students exhibit a decline in performance compared to their baseline test results. However, high-performing students who were paired with low-performing students during the group phase mitigate some of their performance decline upon returning to individual work. This suggests that low-performing students are the primary beneficiaries of collaborative work, deriving lasting benefits from the group experience.²³

²¹The statistically significant differences in mean performance across these three group types, as confirmed by Tukey's HSD test (refer to [Table A2.4](#) in the Appendix.), reinforce our earlier findings that the abilities of both students affect group performance outcomes.

²²We repeat this analysis excluding the individuals suffering bullying, in case they may be driving the results (they improve in group due to social effect bla bla something like that). The results can be seen in [Fig. A3.21](#), showing the same pattern as the analysis with the whole sample.

²³To gain deeper insights into the specific individuals driving these results, we provide summary statistics of the main variables of interest for low-ability students who improved their performance during the group and final tests, as well as high-ability students who experienced a performance decline in these phases. This information is detailed in [Table A2.26](#) in the appendix. Additionally, we conducted t-tests on these variables to compare these individuals with the rest of the sample. The results of these comparisons are presented in [Table A2.27](#) in the appendix. The evidence suggests that high-ability students whose performance declines during the group and final tests tend to be more socially connected (i.e., they are selected more often as friends and select more individuals as friends). The opposite is true for low-ability students who improve their performance during the group phase, who appear to be more socially marginalized. However, no significant differences are observed between low-ability students who improve their final phase scores and those who do not.

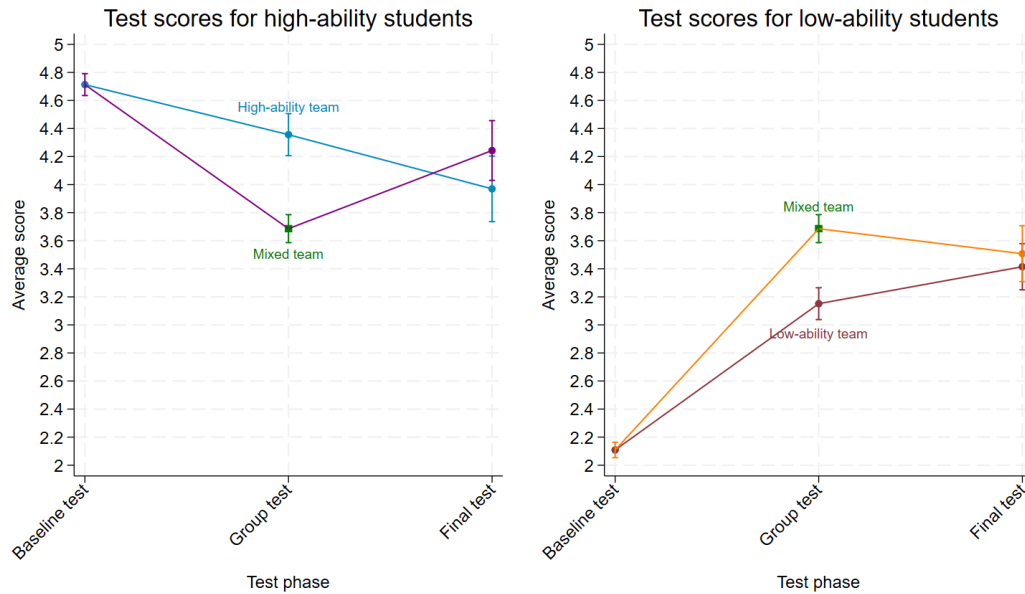


Fig. 3.2. Math average score, depending on the initial ability. Y axis has been adjusted for easier interpretation; scores go from 0 to 10.

Given the proximity of the four average scores for the final test, we conducted a Tukey's HSD test to assess the statistical significance of differences in performance. The results, detailed in [Table 3.6](#), reveal no significant divergences in final test scores for low-performing students, regardless of their partner type during the group phase. Similarly, for high-performing students, the partner during the group phase does not significantly impact their average performance in the final test.

In summary, low-performing students achieve a lower average test score in the final phase compared to high-performing students, with this difference ranging from -0.56 to 0.85 points. However, these scores do not vary significantly based on the group phase partner. The scores obtained by low-performing students in the final phase are significantly higher than their initial individual phase scores, while high-performing students' final phase scores are significantly lower compared to their initial individual scores, indicating a decline in performance from the baseline.

Table 3.6: Tukey's Honestly Significant Difference multiple comparisons for the final test by ability

Student type I	Student type J	Mean	Std. Error	<i>p</i> - value	95% Confidence Interval	
		Difference (I - J)			Lower Bound	Upper Bound
2 low-performing	2 high-performing	0.566***	0.142	0.000	-0.932	-0.200
	Low-high-perform.	0.125	0.136	0.792	-0.223	0.474
2 high-performing	High-low-perform.	0.278	0.154	0.269	-0.117	0.674
	Low-high-perform.	-0.441**	0.154	0.022	-0.837	-0.045
High-low-perform.	2 low-performing	0.844***	0.135	0.000	0.496	1.192
	Low-high-perform.	0.719***	0.148	0.000	-1.098	-0.340

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The second heterogeneity analysis examines the potential effects of the friendship relation between students on performance. As this is one of the primary focuses of the paper, we conduct an in-depth analysis of how student relationships influence outcomes across each test phase, considering whether these effects differ by students' ability levels. Specifically, we explore the impact of friendship separately for high-performing and low-performing students to account for potential variation in outcomes based on initial ability. The results are presented in [Figure 3.3](#) and [Figure 3.4](#).²⁴ The analysis of the difference in mean performance between groups with different levels of friendship relation does not reveal any significant differences in any test phase. Detailed results showing the differences in these averages are provided in [Table A2.5](#) in the Appendix, confirming no significant differences in mean performance among any of the groups. While the ANOVA analysis reveals no significant differences in mean group performance across relationship types, the OLS results highlight the role of interpersonal relationships in shaping group outcomes. The lack of significant findings in ANOVA may reflect its inability to account for individual factors, such as the ability level of the student and their partner (measured by their performance in the baseline test), which are critical determinants of group performance. By controlling for these factors, the OLS analysis uncovers statistically significant negative effects of mutual friendships and enmities on group performance.

²⁴Given the limited sample size of 68 observations in which both students in the same group identified each other as enemies, we have opted to analyze these cases jointly with instances where only one student classified the other as an enemy. Following the same argument, we have opted to analyze the 57 cases in which one individual selected their partner as a friend and the partner selected the individual as an enemy jointly with the neutral case. Separate analyses revealed no substantive differences in outcomes between these two categories but it resulted in substantially higher standard errors.

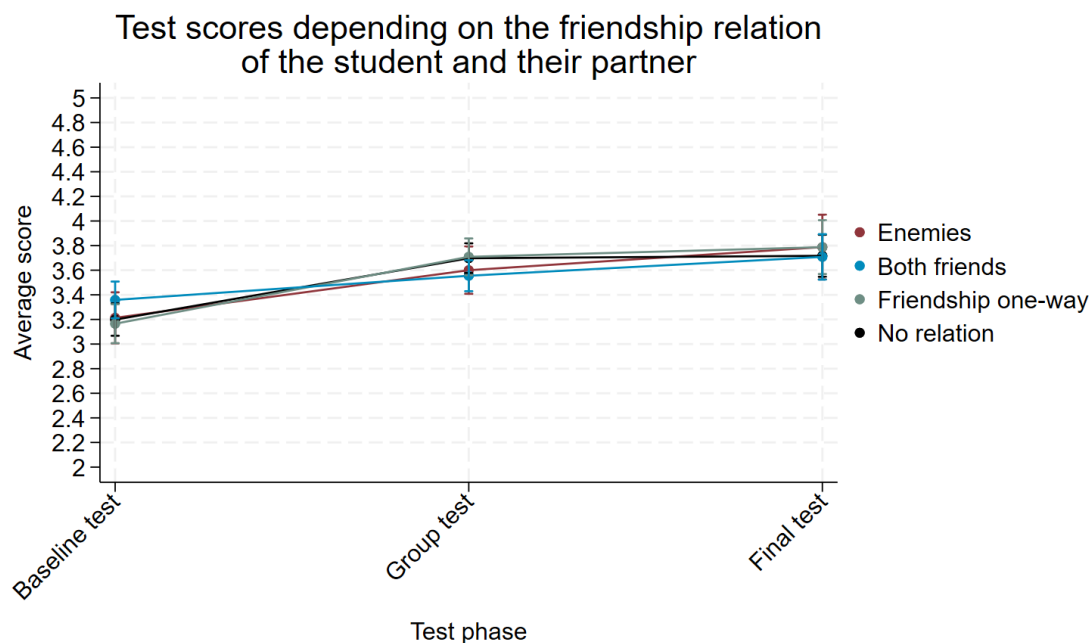


Fig. 3.3. Math average score, depending on the friendship relation between students in the same group.

When incorporating the interaction with ability, the visual analysis becomes more complex, revealing a potential difference for groups of low-ability students composed of two mutual friends and the other types of relationships. The ANOVA results provide weak evidence of a difference in group performance during the group test phase for low-ability students, and when applying Tukey's HSD test, we observe significant differences in mean performance between groups composed of low-ability students where both members mutually selected each other as friends and those where only one member selected the other (i.e., one-sided friendships), as shown in the graph. Notably, no significant differences are found among the remaining groups. Detailed results are presented in [Table A2.6](#), [Table A2.7](#), and [Table A2.8](#) in the Appendix. These findings align partially with the results of the OLS analysis, which identified a negative relationship between group performance and mutual friendships. Specifically, having both individuals in a group mutually identify as friends appears to exert a detrimental impact on performance outcomes.

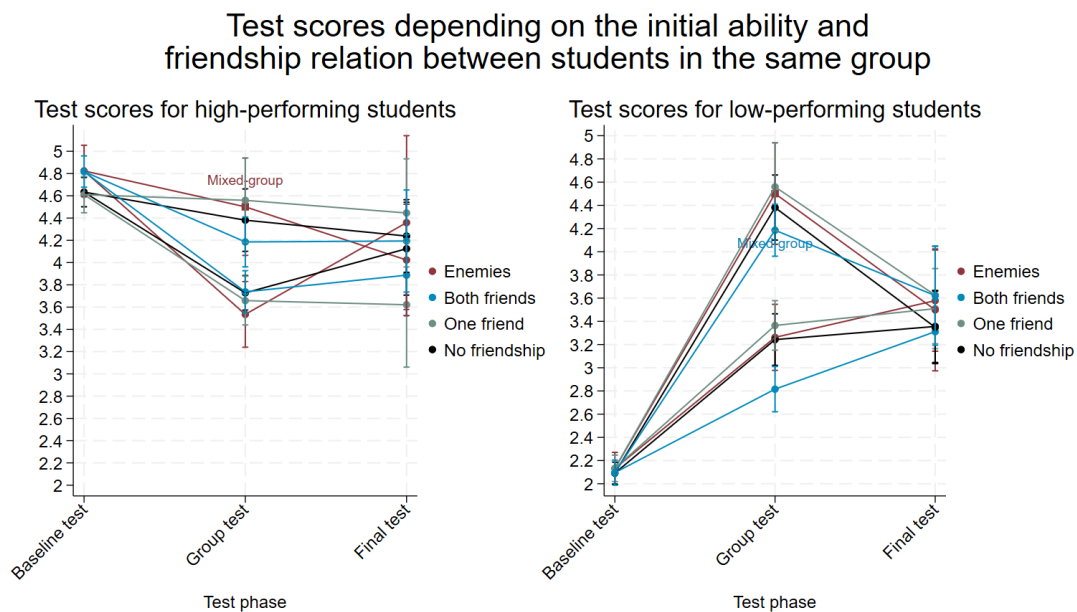


Fig. 3.4. Math average score, depending on the initial ability and friendship relation between students in the same group.

The third heterogeneity analysis examines the potential effects of gender composition. We analyze how students' performance varies across each test phase based on their gender. Additionally, as in the ability analysis, we investigate the combined effect of the student's and their partner's gender across all test phases.

Consistent with our main findings, we observe that female students perform worse than male students on both the individual baseline and final tests, with no significant gender differences in scores during the group test. This is shown in [Figure 3.5](#).²⁵

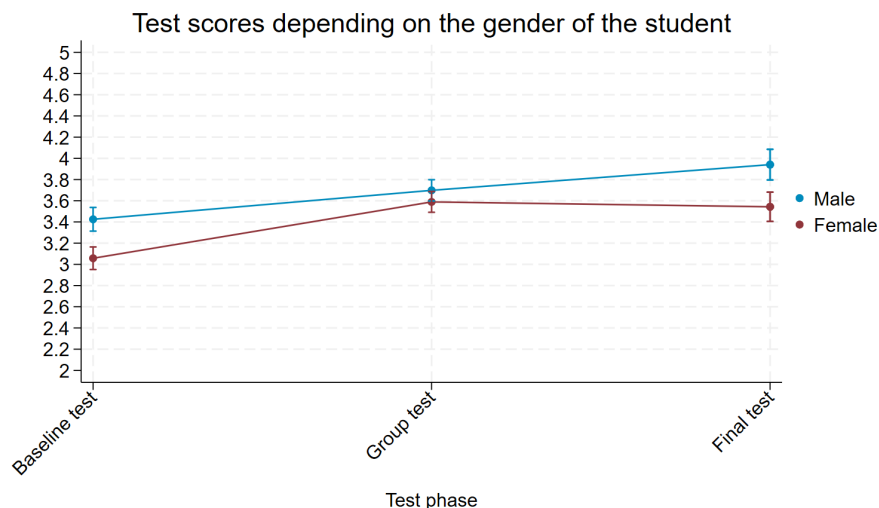


Fig. 3.5. Math average score, depending on the gender of the student.

²⁵Detailed results from the t-tests for differences in means are presented in [Table A2.9](#) in the Appendix.

In [Figure 3.6](#), we incorporate the effect of the partner's gender into the analysis. We observe differences in group phase scores across gender combinations. Tukey's HSD test results, available in [Table 3.7](#), indicate that teams composed of two female students perform worse than those with at least one male student. Specifically, they present a difference in their average score of approximately 0.25 points lower. Additionally, there is some evidence that all-female teams perform worse than all-male teams, with a difference of around 0.23 points, although this difference is only weakly statistically significant.

Table 3.7: Tukey's Honestly Significant Difference multiple comparisons test for group scores by gender

Group Type I	Group Type J	Mean Difference (I - J)	Std. Error	p - value	Lower Bound	Upper Bound
Mixed Team	Both Male	0.019	0.087	0.973	-0.185	0.223
	Both Female	0.248**	0.086	0.012	0.045	0.451
Both Female	Both Male	-0.229*	0.099	0.054	-0.461	0.003

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In the final phase, we find that female students' performance declines for those who previously worked with a male partner during the group test phase. However, this difference is not statistically significant when compared to female students who worked with another female partner. Overall, due to the increase in standard deviation, the only statistically significant difference observed is between groups composed of two male students and those with two female students, with the former outperforming the latter. Detailed results from Tukey's HSD test are provided in [Table A2.10](#) in the Appendix.

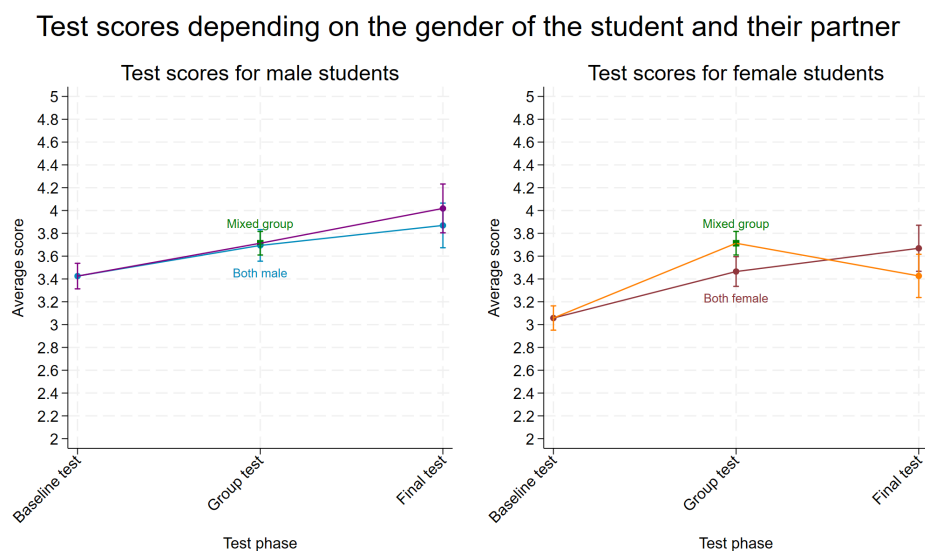


Fig. 3.6. Math average score, depending on the gender of the student and gender of their group partner.

The fourth heterogeneity analysis investigates the potential impact of experiencing bullying²⁶ in the classroom. Specifically, we examine variations in student performance across each test phase, differentiating between students who suffer from bullying and those who do not, as well as considering the bullying status of their assigned partner during the group phase.

Consistent with our main findings, we observe that students experiencing bullying perform worse on both the individual baseline and group tests, with no significant differences on the final test, as shown in Figure 3.7.²⁷ The absence of a statistically significant difference on the final test could be attributed to several factors. First, since the final test was shorter, the negative impacts of bullying on performance may not have been strong enough to be detected. Additionally, it is likely that, after the group work, students who have experienced bullying might benefit from the group dynamics, resulting in performance that aligns more closely with their peers. Further investigation is necessary to understand this dynamic more comprehensively.

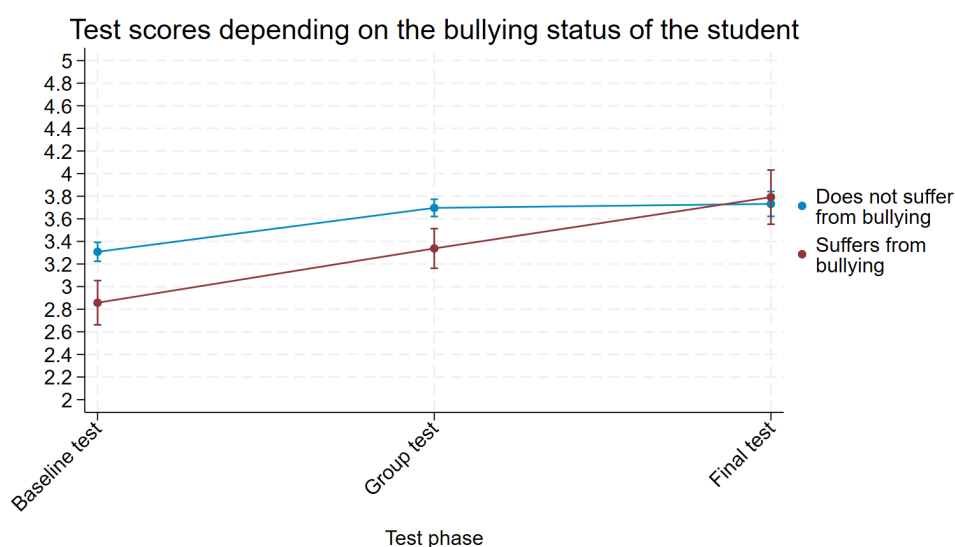


Fig. 3.7. Math average score, depending on the bullying status of the student.

²⁶It is important to remember that the bullying status is self-reported by the students, who can report about themselves and their classmates bullying situation. As shown in the descriptive statistics, around 15% of the students in our sample are potentially suffering from bullying. More information about their characteristics can be found in Table A2.23 in the appendix.

²⁷Detailed results from the t-tests for mean differences are presented in Table A2.11 in the Appendix.

Test scores depending on the bullying status of the student and their partner

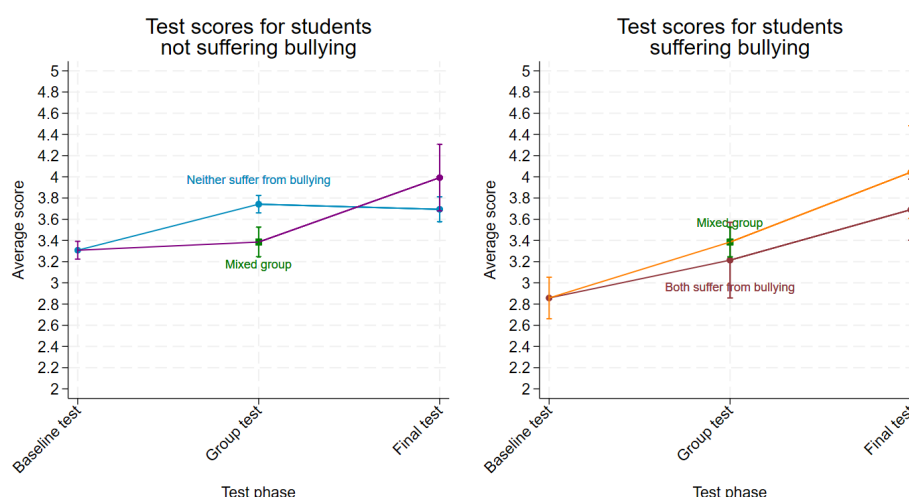


Fig. 3.8. Math average score, depending on the bullying status of the student and their group partner.

When we incorporate the effect of the partner's bullying status into the analysis, as shown in Figure 3.8, we continue to observe no statistically significant differences between groups in the final test.²⁸ However, in the group test, we find that groups in which at least one member experiences bullying perform worse than their counterparts. Specifically, teams in which both students experience bullying score, on average, half a point lower than teams in which neither student experiences bullying. Similarly, teams with only one student affected by bullying score 0.36 points lower than those in which neither student is affected. We find no statistically significant difference in scores between teams in which both students experience bullying and those in which only one does. This suggests that having even one member suffering from bullying in the group is sufficient to significantly reduce the group's overall score relative to other groups. Detailed results are provided in Table 3.8.

Table 3.8: Tukey's Honestly Significant Difference multiple comparisons test for group scores by bullying

Group Type I	Group Type J	Mean Difference (I - J)	Std. Error	p-value	Lower Bound	Upper Bound
Mixed Team	None suffers from bullying	-0.357***	0.087	0.000	-0.561	-0.153
Both suffer from bullying	None suffers from bullying	-0.528***	0.179	0.009	-0.947	-0.108
	Mixed Team	-0.171	0.190	0.641	-0.617	0.275

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

²⁸Results for the ANOVA test are provided in the Appendix, in Table A2.12.

The last heterogeneity analysis examines the potential impact of being born outside of Spain, which might influence test performance due to language barriers or cultural differences.²⁹ We distinguish between students born in Spain and those born abroad, comparing their average scores across all test phases. Graphical results are presented in Figure 3.9.

In line with the baseline analysis, we find no score differences between local and migrant students on the baseline or group tests. However, migrant students perform worse than local students on average in the final test.³⁰ To delve deeper, we also examine the effect of the partner's migration status, as shown in Figure 3.10.

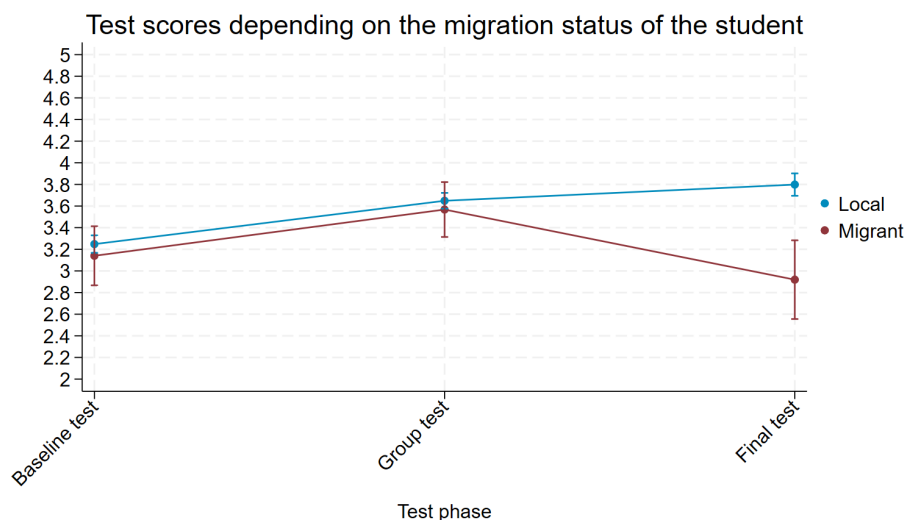


Fig. 3.9. Math average score, depending on the migration status of the student.

When we include the partner's migration status in the analysis, we continue to observe no statistically significant differences between groups in either the individual or group test phases.³¹ However, in the final test phase, migrant students who worked with a local student during the group phase perform worse than local students who had local partners. Given that no other significant differences are observed between the average scores of migrant and local students, this difference appears likely to stem from a spurious relationship.³²

²⁹It is important to notice that only 6.74% of our dataset is composed of students born outside of Spain.

³⁰Detailed results from the t-tests are available in Table A2.13 in the Appendix.

³¹Results for the ANOVA test are available in the Appendix, in Table A2.14.

³²If a genuine effect existed for migrant students, we would expect to see score differences between groups composed of two migrant students and those composed of two local students. Similarly, if the difference resulted from the interaction with a local partner, we would expect a performance difference between migrant students paired with a local partner and migrant students paired with another migrant. However, none of these group averages show statistically significant differences. For detailed results of Tukey's HSD test across all possible pairwise comparisons, please refer to Table A2.15 in the Appendix.

Test scores depending on the migration status of the student and their partner

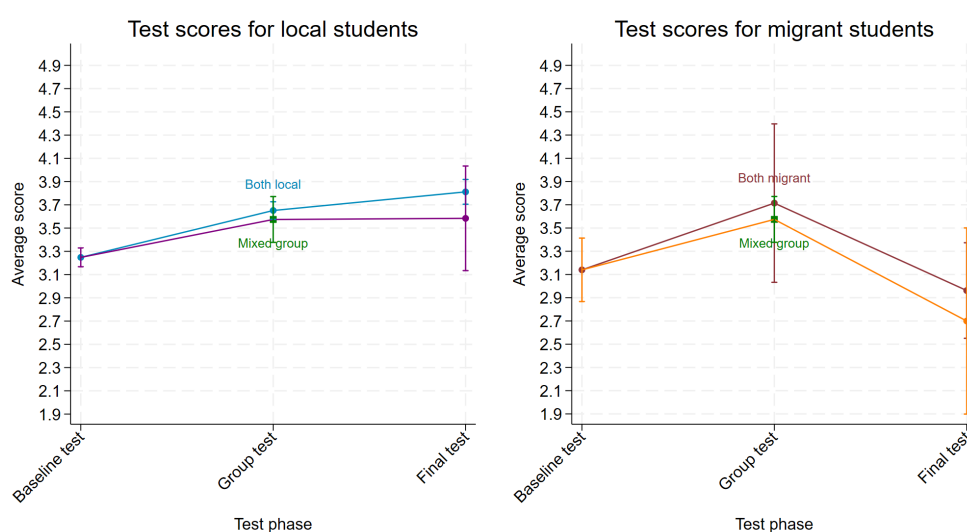


Fig. 3.10. Math average score, depending on the migration status of the student and their group partner.

A2 Well-being effects of the group

When students work in groups, performance is an important metric, although it is not the sole outcome of interest. To expand our analysis, we also examine another critical aspect of teamwork: well-being.

Following the group tests, each student was asked to rate how comfortable³³ they felt working with their assigned partner on a scale from 1 to 6. A screenshot of the specific question they answered is available in the Appendix, in [Figure A3.20](#). Using these responses, we created an indicator categorizing the pairs based on whether both students reported being comfortable, one comfortable and the other uncomfortable, or both uncomfortable. After completing the final individual test, students responded to three questions from the Kidscreen questionnaire concerning their recent social interactions at school.³⁴ From this, we construct an index about their morale state ranging from -3 to 9, which we then use to explore the relationship between comfort level during teamwork and this index. The results of this analysis are shown in [Table 3.9](#).

³³We will use the terms "comfort level" or "well-being" of the student to refer to the same concept.

³⁴The specific questions were: During the last week: 1. Have you felt alone? 2. Have you had fun with your friends? 3. Have you performed well at school?

Table 3.9: Results on morale state

	(1)	(2)
Comfortable	0.621*** (0.173)	0.541*** (0.178)
Female		0.056 (0.207)
Female Partner		-0.150 (0.171)
Score Baseline Test		0.062 (0.056)
Score Baseline Test Partner		0.057 (0.046)
Suffers Bullying		-0.848*** (0.277)
Suffers Bullying Partner		-0.011 (0.236)
Migrant		-0.541 (0.431)
Migrant Partner		-0.452 (0.359)
Class FE	✓	✓
Observations	834	820
Mean of Dep. Variable	6.253	6.290

Standard errors in parentheses, clustered by class. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Our analysis reveals a positive and significant effect of students' reported comfort level during the group phase on their morale state reported at the end of the session. This effect remains robust even when controlling for individual and partner characteristics, with bullying status being the only additional variable that shows a significant impact. These results lead us to interpret the comfort level reported by students as an indicator of their well-being, which turns it into an important factor to consider when forming student groups.

We then proceed to investigate the various factors influencing the level of comfort reported by students. Specifically, we examine how the individual characteristics of each student, the attributes of their assigned partner, and the interaction between these factors, as well as the friendship relation between the students, impact their reported comfort levels. The results of the analyses regarding students' characteristics are presented in [Figure 3.11](#) below.

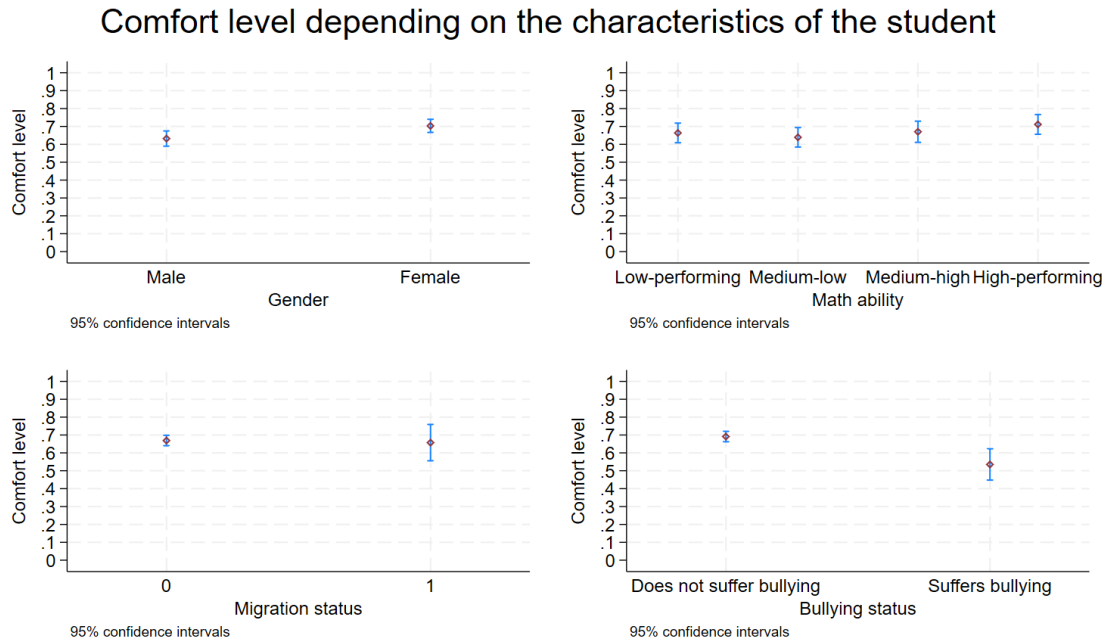


Fig. 3.11. Comfort level depending on the characteristics of the student

We find that the main effect on reported comfort levels is associated with bullying status: students who experience bullying report comfort levels that are 22% lower than those of students who do not experience bullying. Regarding gender, female students report comfort levels approximately 8.5% higher than male students. Nevertheless, this difference is much smaller than the effect of bullying and is only statistically significant at the 5% level. As for ability, we observe a slight U-shaped pattern in comfort levels across different ability levels, even though these differences are not statistically significant. Lastly, with respect to migration status, we find no significant differences in comfort levels between local and migrant students. Detailed results are presented in [Table 3.10](#).

Table 3.10: Regression results on individual characteristics on comfort level

	(1)	(2)	(3)	(4)	(5)
Female	0.0619** (0.0247)				0.0565** (0.0258)
Ability		0.0167 (0.0141)			0.0168 (0.0142)
Migrant			-0.0311 (0.0581)		-0.0289 (0.0576)
Suffers Bullying				-0.158*** (0.0555)	-0.149*** (0.0557)
Class FE	✓	✓	✓	✓	✓
Observations	1628	1628	1628	1628	1628
Mean of Dep. Variable	0.668	0.668	0.668	0.668	0.668

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To explore this relationship further, we extend our analysis to include the specific attributes of each student's assigned partner. We examine the isolated effect of each characteristic independently, as well as their combined impact, aiming to identify how these factors collectively influence reported comfort levels. A visual summary of these findings is provided in Figure 3.12, while Table 3.11 presents detailed results.

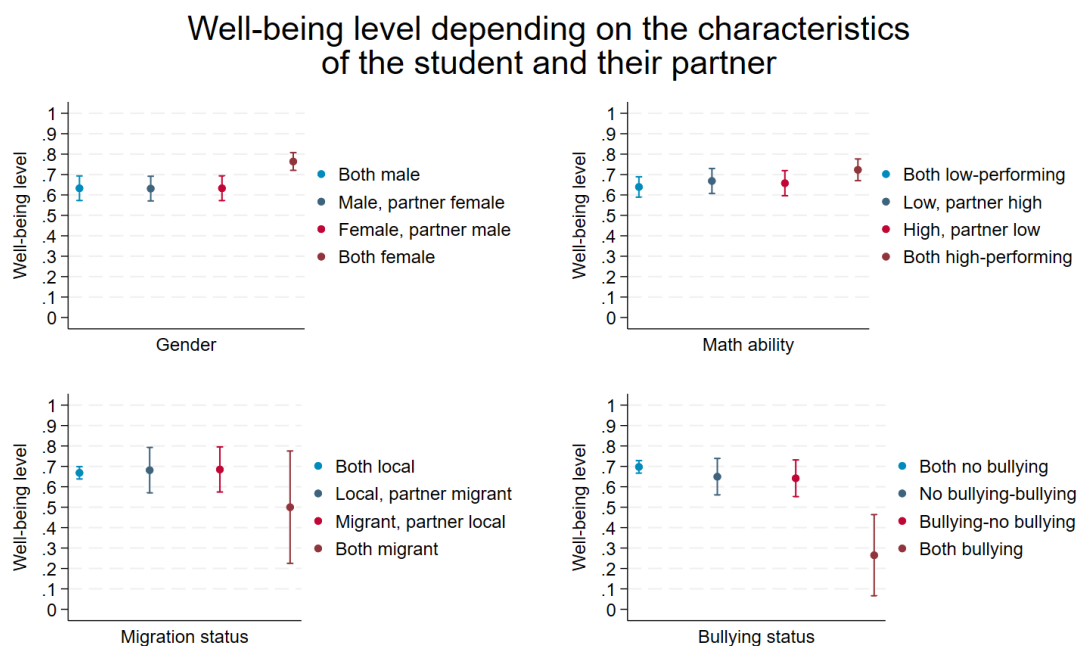


Fig. 3.12. Comfort level depending on the characteristics of the student and their partner

Table 3.11: Regression results of individual and partner characteristics on comfort level

	(1)	(2)	(3)	(4)	(5)
Female	-0.0184 (0.0524)				-0.0382 (0.0513)
Female partner	-0.0195 (0.0524)				-0.0378 (0.0513)
Female $\times$ Female partner	0.163* (0.0871)				0.190** (0.0896)
Ability		-0.0185 (0.0478)			-0.00798 (0.0468)
Ability partner		-0.0140 (0.0479)			-0.00345 (0.0468)
Ability $\times$ ability partner		0.0150 (0.0185)			0.0113 (0.0183)
Migrant			-0.00703 (0.0649)		-0.000674 (0.0662)
Migrant partner			-0.00709 (0.0655)		-0.000143 (0.0670)
Migrant $\times$ Migrant partner			-0.175 (0.200)		-0.196 (0.204)
Bullying				-0.0762 (0.0591)	-0.0580 (0.0588)
Bullying partner				-0.0702 (0.0583)	-0.0512 (0.0579)
Bullying $\times$ Bullying partner				-0.292* (0.169)	-0.318* (0.172)
Class FE	✓	✓	✓	✓	✓
Observations	1628	1628	1628	1628	1628
Mean of Dep. Variable	0.668	0.668	0.668	0.668	0.668

Standard errors in parentheses, clustered by class. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Our analysis reveals that female students' comfort levels are significantly influenced when paired with another female student. Groups comprised of two female students report a comfort level approximately 28% higher than the average, with this effect being statistically significant at the 5% level. Conversely, individuals in all-male or mixed-gender groups do not report significantly different comfort levels. Additionally, when both students in a group have experienced bullying, they report a comfort level about 48% lower than the average, though this effect is only significant at the 10% level.³⁵ No other groups show

³⁵It is important to notice that there are only 68 observations where both students paired together suffer from bullying.

statistically significant differences in reported comfort levels. Groups composed of two high-performing students exhibit a slightly higher comfort level than other groups, but this difference is not statistically significant. Similarly, groups composed of two migrant students report a comfort level that is slightly below the average, though this difference also lacks statistical significance at conventional confidence levels.³⁶

The final aspect we incorporate into our analysis is the students' friendship relations. We examine this at two levels: first, we assess how comfort levels are affected by working with someone based on their friendship relation; second, we analyze the impact on comfort levels depending on the students' group preferences regarding their assigned partner.³⁷ As expected, we observe an upward trend in well-being when students work in groups with peers with whom they have a positive rapport, as shown in Figure 3.13.

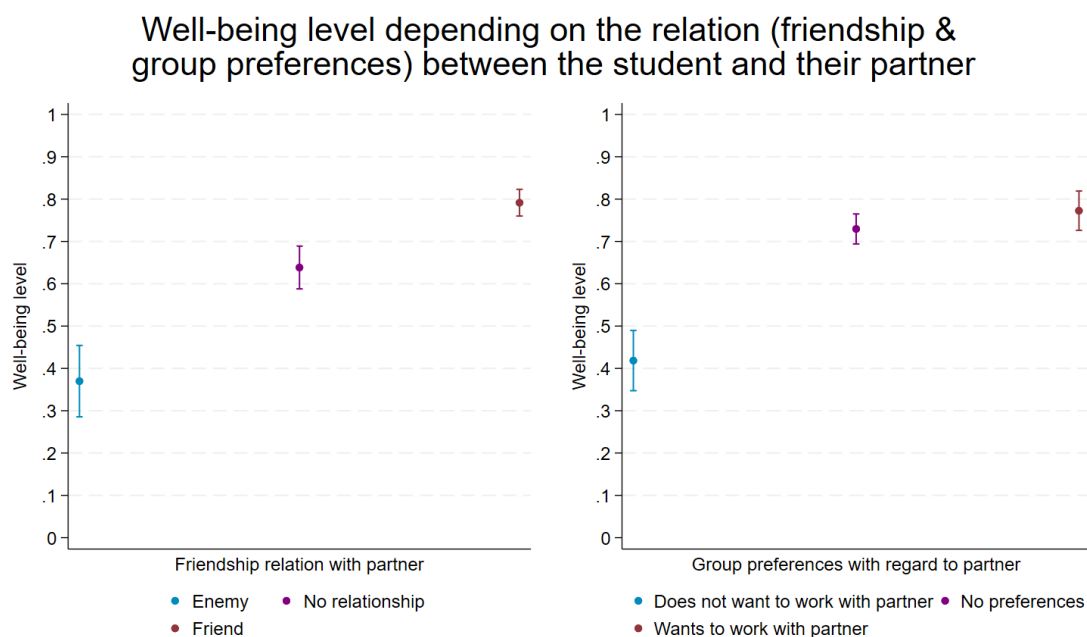


Fig. 3.13. Comfort level depending on the relationship of the student and their partner

Detailed analysis are shown in Table 3.12, revealing that working with an enemy negatively impacts comfort levels by 0.265 points, representing an approximately 39.67% reduction from the mean comfort level. Conversely, working with a friend enhances comfort by 0.177 points, equating to an approximate 26.5% increase from the mean level. A similar pattern emerges for group preferences: being paired with someone a student did not want in their group has a negative effect of 0.31 points, amounting to roughly 46.4% of the mean comfort level. However, collaborating with a desired teammate shows no significant impact compared to working with someone with neutral preference regarding group membership. This finding is consistent with the accompanying graph, which illustrates that the only significantly different mean is found among groups where students were paired with a partner they did not want in their group,

³⁶The t-test results for ability and migration status are available in the Appendix, Table A2.16, showing no significant differences among groups.

³⁷It is important to note that friendship and group preference indicators overlap for nearly 60% of our sample, which suggests that they capture related factors that influence students' well-being when working together. Detailed information is shown in Table A2.24 in the appendix.

with no significant differences in comfort levels between those working with a partner they wanted in their group and those with no specific preference. Detailed results from Tukey's HSD test for these indicators are provided in the Appendix, in [Table A2.17](#) and [Table A2.18](#).

Table 3.12: Regression results of relationship on comfort level

	(1)	(2)
Worked with an enemy	-0.265*** (0.0806)	
Worked with a friend	0.177*** (0.0479)	
Worked with undesired teammate		-0.310*** (0.0676)
Worked with desired teammate		0.0741 (0.0461)
Class FE	✓	✓
Observations	1628	1628
Mean of Dep. Variable	0.668	0.668

Standard errors in parentheses, clustered by class.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To deepen this analysis, we extend our approach to examine the combined effect of both the relationship a student reports and the relationship reported by their partner. This allows us to capture a more comprehensive picture of how mutual and asymmetric perceptions of relationships influence performance. The results, presented in [Table A2.19](#), indicate that working with a partner the student selected as a friend increases the reported comfort level by 0.100 points—an approximate 15% increase from the mean comfort level of the neutral group (where neither student selected each other as a friend or enemy). Notably, if the partner also selected the student as a friend (i.e., the friendship is mutual), the effect remains nearly the same. This suggests that the partner's reciprocal selection does not contribute additional comfort beyond the student's own perception of friendship. ³⁸

Conversely, being paired with a partner the student selected as an enemy leads to a 0.266-point decrease in comfort, corresponding to an approximate 39.82% decline from the mean level—very similar to the findings from the previous analysis. When the enmity is reciprocal (i.e., both students selected each other as enemies), the reduction in reported comfort increases to 0.366 points, representing a 54.8% decline compared to the neutral group. This suggests that when enmity is mutual, the negative effect on comfort intensifies, leading to an additional 0.100-point reduction (approximately 15%) in reported comfort beyond the impact of one-sided enmity alone.

³⁸It is important to note that the results for mutual relationships (both selecting each other as friends or enemies) are only significant at the 5% level, likely due to the smaller number of pairs that meet this criterion.

Table 3.13: Regression results of relationship on comfort level

	(1)	(2)
Friendship both ways	0.0999** (0.0423)	
Friendship one-way	0.100*** (0.0360)	
Enmity one-way	-0.266*** (0.0746)	
Enmity both ways	-0.366** (0.148)	
Mutual group selection		0.134** (0.0511)
One-sided group selection		0.0471 (0.0497)
One-sided no-group selection		-0.249*** (0.0628)
Mutual no-group selection		-0.625*** (0.170)
Class FE	✓	✓
Observations	1625	1625
Mean of Dep. Variable	0.668	0.668

Standard errors in parentheses, clustered by class.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.6 Summary and Concluding Remarks

In this paper, we examine the effects of team composition on academic performance and student well-being. By analyzing a dataset of randomly-created student pairs, we assess the impact of partner characteristics—including ability level and friendship relations—on individual and group outcomes. Our study builds on existing literature addressing the role of peer effects and friendships in collaborative settings, extending this line of inquiry by considering well-being as an educational outcome and situating the research specifically within a school setting, which allows us to offer insights that are directly relevant to school environments. Through our analysis, we find evidence suggesting that group assignments based on academic ability can yield distinct benefits and drawbacks depending on the pairing’s composition, affecting both academic

performance and student well-being. High-ability students may experience different academic outcomes than low-ability students, while friendship-based pairings seem to influence comfort and performance when working in groups in opposite directions. These findings suggest that the effects of group work extend beyond learning gains to include students' social experiences.

Our findings reveal a fundamental trade-off between performance and well-being in group work. On one hand, we confirm that low-performing students experience measurable academic benefits from working in teams, irrespective of their partner's ability level. This suggests that collaborative learning environments can be particularly advantageous for students with lower initial ability, possibly due to the support and motivation that group work provides. Conversely, high-performing students tend to experience a decline in performance when working in groups, aligning with research suggesting that group settings can constrain individual productivity when coordination challenges arise, such as in [Tudge \(1992\)](#) and [Webb \(1982\)](#). The distinction between high-performing and low-performing outcomes indicates that while group work is effective in raising the performance of lower-ability students, it may do so at the expense of those with greater prior achievement. This finding has meaningful implications for educators considering how to structure group work to optimize collective learning gains.

On the other hand, our analysis reveals that pre-existing social relationships between group members significantly influence both performance and well-being. Specifically, we find that friendship relations among students foster reported comfort levels in group settings, which we use as a proxy for well-being in the classroom, underscoring the social benefits of familiar partnerships. This indicates that specific pairing strategies in group assignments can enhance non-cognitive outcomes such as morale. However, this increased comfort does not translate into better academic outcomes. Instead, mutual friendships within a group lead to a decline in performance, likely due to distractions, a more relaxed approach to the task, or reduced pressure to engage critically with the material. Similarly, students paired with mutual enemies also exhibit a significant drop in performance, likely due to communication breakdowns or conflict avoidance. Overall, these findings suggest that both strong positive and negative pre-existing relationships between group members can hinder academic performance. Thus, our results highlight the necessity of balancing the effects on students' well-being with the pursuit of academic gains when evaluating the efficacy of group work.

This trade-off between well-being and performance, along with our broader findings, indicate several potential implications for educational policies and classroom practices. First, pairing students based solely on academic ability could yield optimal performance outcomes; however, the potential trade-offs for high-ability students should be weighed carefully. Policies that aim to boost the performance of lower-ability students without impeding higher-ability students may consider differentiated grouping strategies, balancing both cognitive and social benefits. Moreover, by including measures of student well-being, we emphasize that policies should address both academic and social dimensions. For example, settings focused on pairing students based on social compatibility could be instrumental in improving classroom cohesion and morale, particularly for those students who may feel marginalized or isolated. Furthermore, pairing strategies that mitigate the negative effects of both friendships and enmities—such as rotating partners over time or incorporating structured collaboration guidelines—could help optimize learning while preserving social cohesion.

This study leaves space for further research on aspects that highlight its limitations. For example, one constraint is the potential for confounding variables not captured in the data, such as students' personal goals or learning styles. Future research could investigate whether specific types of tasks benefit more from friendship-based groupings or if certain characteristics (e.g., communication skills or teamwork ability) influence the observed effects. Another limitation stems from our use of the comfort variable,

The Cost of Friendship: Who Is a Better Working Partner?

which, while offering insight into students' subjective experience, may not fully encompass all aspects of well-being. Expanding the range of well-being indicators, such as stress levels, resilience, and emotional development, could provide a more comprehensive perspective on the effects of peer interactions on students' overall well-being.

3.7 Appendix

A1 Robustness checks

Randomization checks

To verify the effectiveness of the randomization process in our experiment, we conducted a series of tests. Specifically, we divided our sample into two subsets: one containing the observations for the first individual of each group, and the other for the second individual. Using this partitioned dataset, we compared the sample means and variances across key variables: baseline score, group score, and final score. [Table A2.21](#) reveals no significant differences in mean test scores between the two members of the groups, and [Table A2.22](#) confirms the same for variance. These findings indicate that the data generation process does not differ between students and their partners. Consequently, this suggests that the randomization process used to form student groups was effective, as no significant differences were observed in these statistical measures.

Separating students using their performance in a logic test

During the experimental session, after completing the initial individual math test but before engaging in the group math test with their assigned partner, students were asked to solve a logic test based on Raven's Matrices individually. In this section, we use the results of this logic test to classify students into high-ability and low-ability, based on whether their performance is above or below the median. Using this reclassification, we replicate the analysis of their performance across the three math tests, now based on their own logic ability and that of their partner. The results of this analysis are presented in [Figure A3.22](#).

In this classification, all students benefit from group work. Low-ability students improve their performance even when paired with another low-ability student, and even more so when paired with a high-ability student—although this improvement is less pronounced than when classifying students based on their initial math performance. High-ability students, on the other hand, show performance gains when paired with another high-ability student. Unlike in the math-based classification, when paired with a low-ability student, their performance does not decline but rather remains statistically indistinguishable from their individual performance. No differences are observed on the final individual test when using logic compared to when using math performance to classify students.

Notably, 43% of students switch categories between the math-based and logic-based classifications (i.e., some students classified as high-ability in math are classified as low-ability in logic, and vice versa). This suggests that performance in one domain does not fully predict performance in the other. Some students who performed poorly in math may still possess strong reasoning abilities that enhance their capacity to engage in collaborative problem-solving, even if they struggle with calculations. If logic-based classification better reflects how students approach problems, then even students classified as low-ability in logic may still actively contribute during group discussions. This could foster collective problem-solving strategies that enhance overall performance.

Thus, logic-based classification may create a more balanced group structure, where all students benefit from the nature of the task itself. This suggests that effective collaboration depends not only on ability gaps but also on how students engage with problems together.

Performance analysis separating by initial performance and gender

In analyzing the impact of gender on team performance, we find that teams composed of two female students perform worse than those formed by two male students or mixed-gender teams. To further investigate this pattern, we conduct a combined analysis of gender and ability, but using their scores on the logic test instead of the math test to classify students, in order to add robustness to the results. They are presented in [Figure A3.23](#) for male students and [Figure A3.24](#) for female students.

For high-ability male students, teams composed of two high-ability males exhibit marginally higher performance than those consisting of a high-ability male and a high-ability female. In turn, these mixed teams perform better than those composed of a high-ability and a low-ability student, regardless of gender. However, none of these differences are statistically significant at conventional levels, as shown in [Table A2.27](#).

Low-ability male students in all-male teams underperform relative to mixed-gender pairs (regardless of the female partner's ability). However, none of these differences reach statistical significance. The only significant difference emerges between teams composed of a low-ability male and a high-ability male, who perform significantly better than teams formed by two low-ability male students. The results of the Tukey test for mean differences are reported in [Table A2.28](#).

For high-ability female students, we find no significant performance differences when they are paired with a low-ability student, whether male or female. Their performance improves when partnered with another high-ability female student and increases further when paired with a high-ability male student, though these differences remain statistically insignificant. The only statistically significant difference is observed in teams consisting of a high-ability male and a high-ability female, who perform better than teams composed of a high-ability female and a low-ability student, irrespective of the latter's gender. Additionally, the average performance of high-ability female students remains lower than that of high-ability male students, consistent with the findings from our gender analysis. This suggests that high-ability female students may experience a greater decline in performance when working with low-ability partners, potentially due to increased efforts to assist their teammates in understanding the test problems, which may come at the expense of their own performance.

High-ability female students show no performance differences when paired with low-ability peers (male or female). Their performance improves modestly when partnered with another high-ability female and increases further when paired with a high-ability male student, although neither difference attains statistical significance. The only statistically significant difference is observed in teams consisting of a high-ability male and a high-ability female, who perform better than teams composed of a high-ability female and a low-ability student, irrespective of the latter's gender. Additionally, the average performance of high-ability female students remains lower than that of high-ability male students, consistent with the findings from our gender analysis. This suggests that high-ability female students may experience a greater decline in performance when working with low-ability partners, potentially due to increased efforts to assist their teammates in understanding the test problems, which may come at the expense of their own performance.

For low-ability female students, a similar pattern emerges. Teams composed of two low-ability female students perform worse than those formed by a low-ability female student and a high-ability female student. Performance is similar between teams formed by a low-ability female and a low-ability male student and a low-ability and a high-ability female, though both perform worse than teams composed of a low-ability female and a high-ability male. However, the only statistically significant difference is between teams of two low-ability female students and those of a low-ability female and a high-ability male student. The results of the Tukey test for mean differences are presented in [Table A2.29](#).

Overall, these findings suggest a dual effect for low-ability female students. Compared to being paired with another low-ability female student, their performance improves by a similar magnitude whether they are paired with a male student of the same ability level or with a high-ability student of the same gender, indicating that both the gender and ability of their partner contribute to performance gains, with male partners potentially exerting an additive effect.

A2 Figures

Selecciona els teus amics/amigues. Recorda que pots seleccionar tots els companys del teu curs.

2º ESO - B 2º ESO - A 1º ESO - B 1º ESO - A

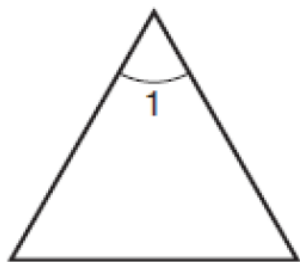
☐ Elijah Ross - 2º ESO A	☐ William Mitchell - 2º ESO A
☐ Sophia Thompson - 2º ESO A	☐ Elizabeth Reed - 2º ESO A
☐ Noah Bennett - 2º ESO A	☐ Michael Nelson - 2º ESO A
☐ Isabella Allen - 2º ESO A	☐ Grace Mitchell - 2º ESO A
☐ Lucas Baker - 2º ESO A	☐ Daniel Parker - 2º ESO A
☐ Olivia Jenkins - 2º ESO A	☐ Emily Green - 2º ESO A
☐ Liam Wood - 2º ESO A	☐ Samuel Phillips - 2º ESO A
☐ Ava Martinez - 2º ESO A	☐ Scarlett Wilson - 2º ESO A
☐ Benjamin Turner - 2º ESO A	☐ James Carter - 2º ESO A

Fig. A3.1: Example of the screen students were shown when asked to select their classmates. Question asked changed, asking to select friends/people they do not have a good relationship with/people they want/do not want in their team. Names are fictional.

Selecciona els teus millors amics/amigues

☐ Daniel Parker - 2º ESO A	☐ Henry Powell - 2º ESO B
☐ Lucas Baker - 2º ESO A	
☐ Isabella Allen - 2º ESO A	
☐ Chloe Patterson - 2º ESO B	

Fig. A3.2: Example of the screen shown directly after the previous one, with a subsample of the list showing only those selected in the previous case, and asking now to select from this subsample their best friends/people they like the least/most important people they want/do not want in their team. Names are fictional.



What is the size of angle 1?

Quant mesura l'angle 1 del triangle equilàter de la imatge?

- ☐ | 90°
- ☐ | 45°
- ☐ | 30°
- ☐ | 60°

Fig. A3.3: Example of a math question for a secondary school student.

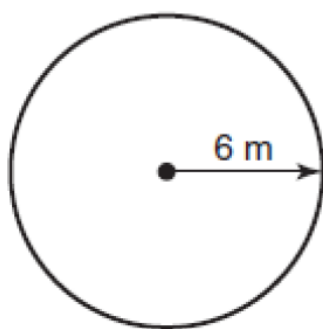
$$\frac{9}{12} \times \left(1 + \frac{1}{4}\right) = ?$$

What is the result?

A què és igual aquesta expressió?

- ☐ | 24/5
- ☐ | 1/88
- ☐ | 15/16
- ☐ | 12/48

Fig. A3.4: Example of a math question for a secondary school student.



What is the area of the circle?

Quina és l'àrea d'aquest cercle?

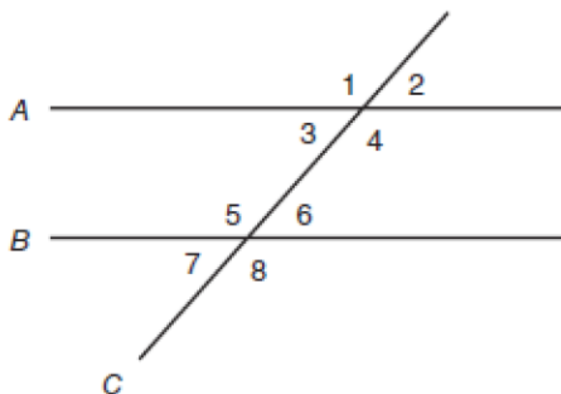
☐ | 97,02 metres quadrats

☐ | 18,84 metres quadrats

☐ | 452,16 metres quadrats

☐ | 113,04 metres quadrats

Fig. A3.5: Example of a math question for a vocational school student.



The lines A and B are parallel and line C crosses both of them. The angle marked as 1 measures 145° . What is the sum of the angles marked as 1, 2, 5, and 6?

Les línies A i B són paral·leles i la línia C creua totes dues. L'angle marcat com a 1 mesura 145° . Quant fa la suma dels angles marcats com a 1, 2, 5 i 6?

☐ | 160°

☐ | 360°

☐ | 240°

☐ | 180°

Fig. A3.6: Example of a math question for a secondary school student.

The Cost of Friendship: Who Is a Better Working Partner?

Quin és el volum d'un cilindre de radi 4 cm i altura 6 cm?

☐ | 75,36 centímetres cúbics

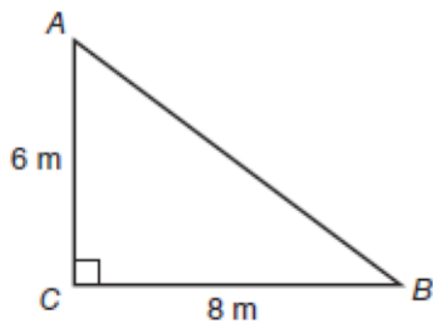
☐ | 150,72 centímetres cúbics

☐ | 301,44 centímetres cúbics

☐ | 602,88 centímetres cúbics

What is the volume of a cylinder with a radius of 4 cm and a height of 6 cm?

Fig. A3.7: Example of a math question for a vocational school student.



Quant mesura el costat AB d'aquest triangle?

☐ | 5 m

☐ | 9 m

☐ | 10 m

☐ | 14 m

What is the length of side AB of this triangle?

Fig. A3.8: Example of a math question for a secondary school student.

$$3a + 4b = 12$$

$$2a + 2b = 15$$

$$a = ?, b = ?$$

What is the solution to this equation?

Quina és la solució de la equació?

- ☐ | a = 12; b = 10
- ☐ | a = 18; b = -10,5
- ☐ | a = -10; b = 12
- ☐ | a = -4; b = 9,5

Fig. A3.9: Example of a math question for a vocational school student.

El professor de català ha assignat 192 pàgines per llegir. Si l'Assumpta comença a llegir el dilluns i llegeix 48 pàgines cada dia, quin dia de la setmana haurà acabat la feina?

- ☐ | Dissabte
- ☐ | Divendres
- ☐ | Dimecres
- ☐ | Dijous

The Catalan teacher has assigned 192 pages to read. If Assumpta starts reading on Monday and reads 48 pages each day, on what day of the week will she have finished the task?

Fig. A3.10: Example of a math question for a secondary school student.

La Patrícia va realitzar un viatge en moto. Va conduir a una velocitat constant de 50 km/h i va trigar 3 hores a arribar a la seva destinació. Quant hauria trigat si hagués conduït a una velocitat de 40 km/h?

- ☐ | 4,25 h
- ☐ | 4 h
- ☐ | 4,45 h
- ☐ | 3,75 h

Patrícia went on a motorcycle trip. She drove at a constant speed of 50 km/h and took 3 hours to reach her destination. How long would it have taken if she had driven at a speed of 40 km/h?

Fig. A3.11: Example of a math question for a secondary school student.

The Cost of Friendship: Who Is a Better Working Partner?

Quant valdrà un article de 200 € després de rebre primer un descompte del 10 % i després un descompte del 15 % adicional?

☐ 75 €
☐ 150 €
☐ 153 €
☐ 175 €

How much will an item priced at €200 be after first receiving a 10% discount and then an additional 15% discount?

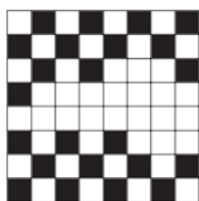
Fig. A3.12: Example of a math question for a secondary school student.

La Felisa té un compte d'estalvis a tipus d'interés fixe. Si en 4 mesos aquest compte li dona un rendiment de 60€, quants en rendirà en 1,5 anys (1 any i mig)?

☐ 320 €
☐ 240 €
☐ 270 €
☐ 200 €

Felisa has a savings account with a fixed interest rate. If in 4 months this account yields €60, how much will it yield in 1.5 years (1 and a half years)?

Fig. A3.13: Example of a math question for a vocational school student.



How many black squares are left to paint on this chessboard?

Quantes caselles negres falta pintar en aquest tauler d'escacs?

☐ 8
☐ 10
☐ 9
☐ 7
☐ 11
☐ 11

Fig. A3.14: Example of a math question for elementary school students.

En Jordi té una capsa amb boles. També té boles fora de la capsa. A cada torn fa el següent: treu la meitat de boles de la capsa, en posa dues a dins i afegeix a la capsa tantes boles com conté ara aquesta. Ho ha fet 35 vegades i la capsa conté finalment 148 boles. Quantes n'hi havia al començament?

☐ | 8

☐ | 78

☐ | 48

☐ | 100

☐ | 113

Jordi has a box with balls. He also has balls outside the box. Each turn he does the following: he takes out half of the balls from the box, puts two inside, and then adds as many balls to the box as it contains now. After doing this 35 times, the box finally contains 148 balls. How many balls were there at the beginning?

Fig. A3.15: Example of a math question for elementary school students.

Hi ha cinc persones en una sala. Algunes d'elles sempre diuen la veritat i les altres sempre menteixen i totes elles saben com és cadascuna. Vam preguntar a cadascuna de les persones de la sala "Quantes persones mentideres hi ha a la sala?". Les respostes van ser 1, 2, 3, 4 i 5. Quantes persones mentideres hi ha realment a la sala?

☐ | 4

☐ | 3

☐ | 1

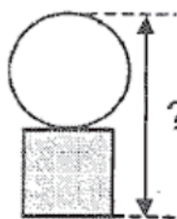
☐ | 2

☐ | 5

☐ | 5

There are five people in a room. Some of them always tell the truth, and others always lie, and each person knows who is who. We asked each person in the room, "How many liars are there in the room?" The answers were 1, 2, 3, 4, and 5. How many liars are actually in the room?

Fig. A3.16: Example of a math question for elementary school students.



Narcís has drawn a lamp composed of a circle and a square at the base (as shown in the drawing). The radius of the circle is 4 cm, and the perimeter of the square is 12 cm. What is the total height of the lamp?

En Narcís ha dibuixat un llum format per un cercle i un quadrat a la base (tal i com es pot veure en el dibuix). El radi del cercle és de 4 cm i el perímetre del quadrat és de 12 cm. Quina és l'altura total del llum?

☐ | 5 cm

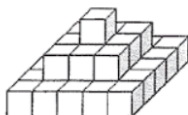
☐ | 7 cm

☐ | 11 cm

☐ | 16 cm

☐ | 20 cm

Fig. A3.17: Example of a math question for elementary school students.



The pyramid in the figure has three levels, each completely filled with small cubes. How many small cubes are there in total?

La piràmide de la figura té tres pisos, cadascun dels quals està tot ple de petits cubs. Quants petits cubs hi ha en total?

☐ | 36

☐ | 46

☐ | 25

☐ | 35

☐ | 20

Fig. A3.18: Example of a math question for elementary school students.

El koala tenia dues branques per dinar. Cada branca tenia 10 fulles. El koala es va menjar algunes fulles de la primera branca. Després, de la segona branca, es va menjar tantes fulles com havia deixat a la primera. Quantes fulles sense menjar van quedar en total a les dues branques?

☐ 10
☐ 8
☐ 6
☐ 15
☐ 5

The koala had two branches for lunch. Each branch had 10 leaves. The koala ate some leaves from the first branch. Then, from the second branch, it ate as many leaves as were left on the first. How many uneaten leaves were left in total on both branches?

Fig. A3.19: Example of a math question for elementary school students.



Indicate how comfortable you felt working with your group partner, whether your experience was satisfactory or not.

Indica com de còmode t'has sentit al treballar amb el teu company de grup, si la teva experiència ha estat satisfactòria o no.

☐ Molt incòmode	Very uncomfortable
☐ Bastant incòmode	Pretty uncomfortable
☐ Una mica incòmode	Somewhat uncomfortable
☐ Una mica còmode	Somewhat comfortable
☐ Bastant còmode	Pretty comfortable
☐ Molt còmode	Very comfortable

Fig. A3.20: Screenshot of the comfortable level question.

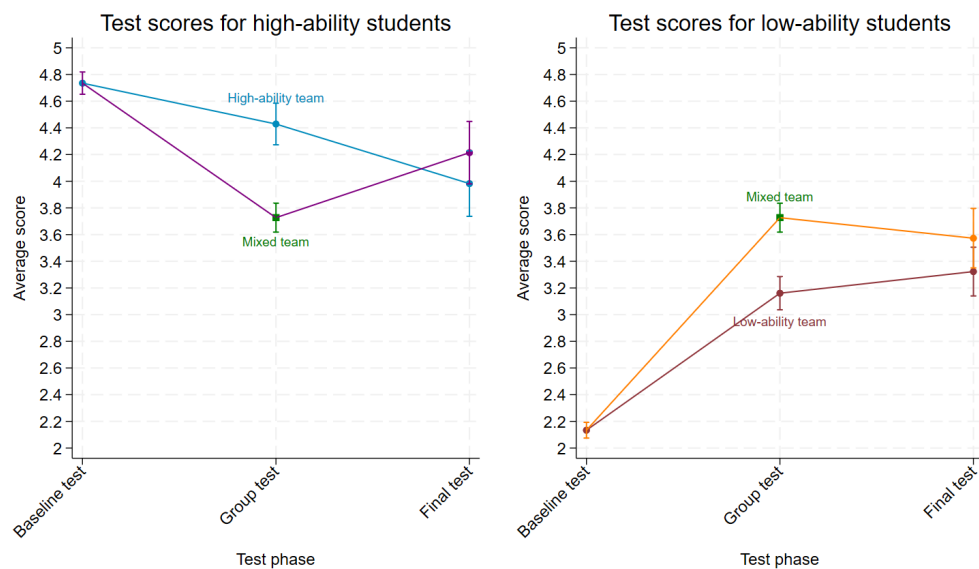


Fig. A3.21: Math average score, depending on the initial ability, excluding students suffering from bullying.

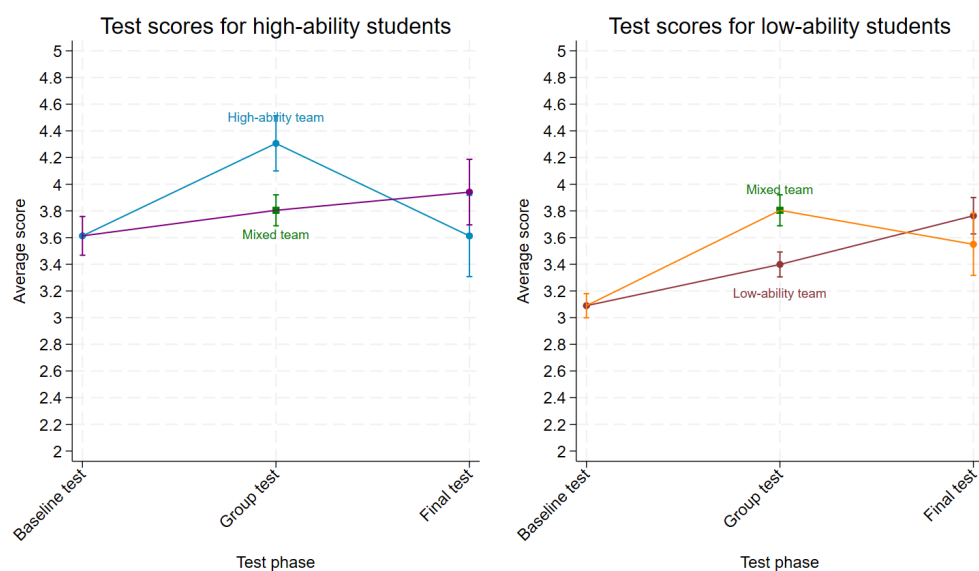


Fig. A3.22. Math average score, depending on the initial logic ability.

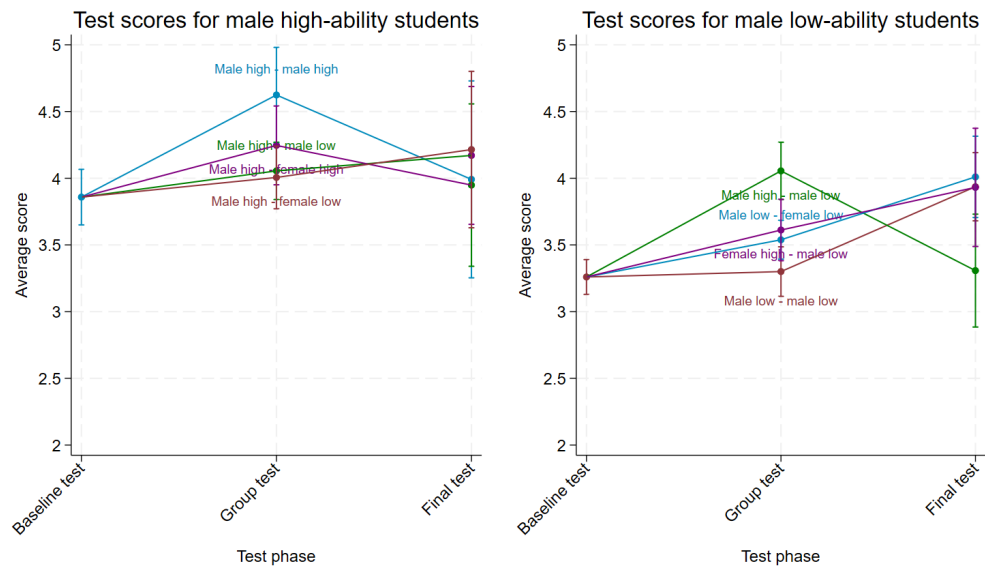


Fig. A3.23. Math average score, depending on the initial logic ability and gender for male students.

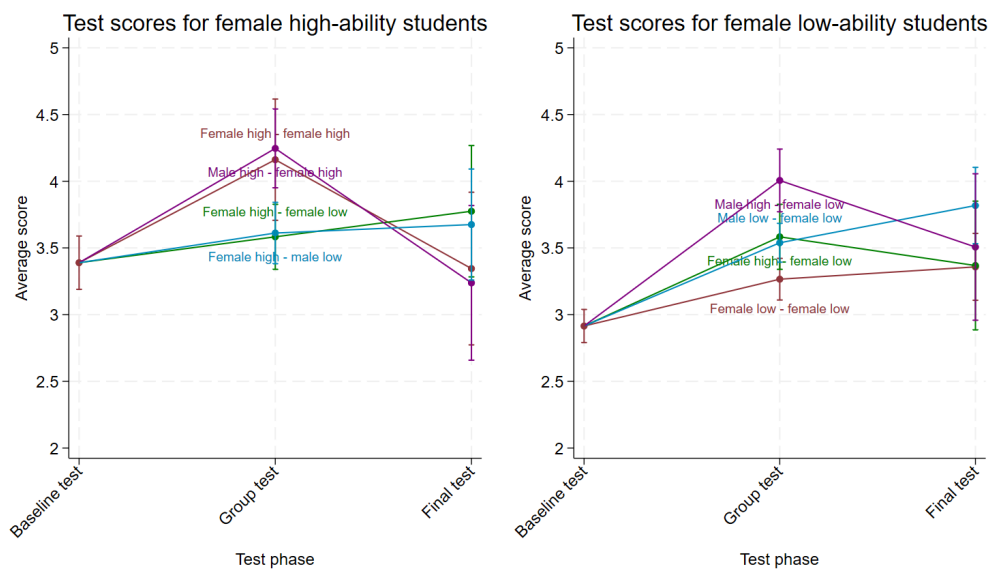


Fig. A3.24. Math average score, depending on the initial logic ability and gender for female students.

A3 Tables

Table A2.1: School characteristics.

School	City	Type of school	Socioeconomic index (IST)	Diff. wrt. avg. (101.8)	% population wrt. low educational attainment	Diff. wrt. avg (14.8)	% foreigners from low or middle-income countries	Diff. wrt. avg. (14.3)	Average income per capita (€)	Diff. wrt. avg. (15871)
1	Badalona	Charter	113.3	11.5	14.3	-0.5	8.3	-6	19821	3950
2	Badalona	Charter	80.5	-21.3	19.3	4.5	23.8	9.5	10901	-4970
3	Barcelona	Public	84.4	-17.4	20.6	5.8	16.2	1.9	18298	2427
4	Barcelona	Charter	99.8	-2.0	14.5	-0.3	14.5	0.2	13427	-2444
5	Barcelona	Public	108.0	6.2	11.6	-3.2	14.1	-0.2	15062	-809
6	Barcelona	Charter	120.8	19.0	6.1	-8.7	9.3	-5.0	20727	4856
7	Barcelona	Public	119.9	18.1	5.2	-9.6	13.6	-0.7	22793	6922
8	Barcelona	Public	118.5	16.7	5.7	-9.1	10.7	-3.6	20403	4532
9	Barcelona	Public	117.1	15.3	11.3	-3.5	7.0	-7.3	18082	2211
10	Barcelona	Charter	103.3	1.5	14.1	-0.7	14.8	0.5	15113	-758
11	Barcelona	Public	110.2	8.4	11.1	-3.7	10.1	-4.2	17342	1471
12	Barcelona	Public	104.0	2.2	13.9	-0.9	14.2	-0.1	15733	-138
13	Ripollet	Public	118.1	16.3	8.1	-6.7	1.8	-12.5	15215	-656
14	Sabadell	Charter	119.5	17.7	6.4	-8.4	1.6	-12.7	14833	-1038
15	Sabadell	Public	100.2	-1.6	15.6	0.8	8.0	-6.3	13715	-2156
16	Sant Quirze del Vallès	Public	128.0	26.2	5.7	-9.1	2.6	-11.7	20806	4935
17	Santa Coloma de Gramanet	Charter	89.4	-12.4	19.3	4.5	13.7	-0.6	11847	-4024
18	Santa Perpètua de Mogoda	Charter	105.1	3.3	13.3	-1.5	11.0	-3.3	15234	-637
19	Viladecavalls	Public	107.6	5.8	13.8	-1.0	1.1	-13.2	14630	-1241
Average			107.77	5.97	12.1	-2.7	10.34	-3.96	16525.37	654.37

Source: Own elaboration using data provided by the [Institut d'Estadística de Catalunya](#).

The difference is with respect to the average for the “Comarca del Barcelonès”.

There are a total of 1166 elementary and high schools in the Barcelonès, with a 46% being public schools and a 54% being charter schools. Compared to our sample, we have 42% charter schools and 58% public schools.

Table A2.2: Regression of individual characteristics and friendship relationship on the performance of the group, complete regression.

Friendship reciprocity	-0.207* (0.109)
Friendship one-way	-0.150 (0.109)
Enemy with friend	0.114 (0.317)
Enemy one-way	-0.0714 (0.157)
Enemy both ways	-0.401* (0.215)
Score baseline test	0.196*** (0.0264)
Score baseline test partner	0.173*** (0.0274)
Female	-0.0126 (0.112)
Female partner	-0.0405 (0.110)
Female $\times$ Female Partner	-0.0194 (0.181)
Suffers bullying	-0.0567 (0.0861)
Partner suffers bullying	-0.0646 (0.0725)
Migrant	-0.109 (0.121)
Migrant partner	-0.0377 (0.130)
Class FE	✓
Observations	1635
Mean of Dep. Variable	3.645

Standard errors in parentheses, clustered by class. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Omitted category is “No relationship between students”.

Table A2.3: Regression of individual characteristics and friendship relationship on the performance on the final phase test.

	(1)	(2)	(3)	(4)	(5)
Friendship relation	0.0403 (0.0582)			0.0235 (0.0586)	0.0220 (0.0578)
Same-sex team		-0.0891 (0.0936)		-0.0501 (0.0962)	
Score baseline test			0.295*** (0.0368)	0.294*** (0.0369)	0.287*** (0.0375)
Score baseline test Partner			0.0536 (0.0381)	0.0521 (0.0381)	0.0537 (0.0385)
Female					-0.200* (0.110)
Female Partner					0.0934 (0.0826)
Wealth					0.0728 (0.0811)
Wealth Partner					-0.0641 (0.0791)
Suffers Bullying					-0.0686 (0.125)
Suffers Bullying Partner					0.198 (0.155)
Migrant					-0.483** (0.195)
Migrant Partner					0.0831 (0.184)
Class FE	✓	✓	✓	✓	✓
Observations	1598	1598	1598	1598	1572
Mean of Dep. Variable	3.741	3.741	3.741	3.741	3.733
Standard errors in parentheses, clustered by class.			* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$		

Table A2.4: Tukey Honestly Significant Difference multiple comparisons for the group test by ability

Student type I	Student type J	Mean Difference (I - J)	Std. Error	p - value	Lower Bound	Upper Bound
Mixed group	2 high-performing	-0.670***	0.091	0.000	-0.884	-0.457
	2 low-performing	0.535***	0.077	0.000	0.353	0.716
2 high-performing	2 low-performing	-1.205***	0.096	0.000	-1.431	-0.979

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the Tukey's HSD test results for the mean differences, standard deviations, p-values, and 95% confidence intervals across all ability pairings in the group test phase.

Table A2.5: ANOVA results for all test phases by students' friendship relation.

Score Type	Observations	Model SS	Residual SS	Total SS	F-value	p-value
Baseline Score	1,640	9.969	4,198.640	4,208.609	1.29	0.275
Group Score	1,638	7.384	3,408.463	3,415.847	1.18	0.316
Final Score	1,598	2.091	6,658.305	6,660.396	0.17	0.919

This table presents the ANOVA results for all friendship pairings in the three test phases. The analysis did not identify statistically significant differences in any of the test phases.

Table A2.6: Low-ability students: ANOVA results for all test phases by students' friendship relation.

Score Type	Observations	Model SS	Residual SS	Total SS	F-value	p-value
Baseline Score	927	0.334	674.601	674.935	0.15	0.928
Group Score	927	14.010*	1784.014	1798.024	2.42	0.065
Final Score	906	6.719	3409.243	3415.963	0.59	0.620

This table presents the ANOVA results for all friendship pairings in the three test phases, specifically for low-performing students.

Table A2.7: High-ability students: ANOVA results for all test phases by students' friendship relation.

Score Type	Observations	Model SS	Residual SS	Total SS	F-value	p-value
Baseline Score	713	6.791	793.531	800.321	2.02	0.109
Group Score	711	0.475	1,440.840	1,441.315	0.08	0.972
Final Score	692	3.084	3,068.189	3,071.273	0.23	0.875

This table presents the ANOVA results for all friendship pairings in the three test phases, specifically for high-performing students.

Table A2.8: Low-ability students: Tukey's Honestly Significant Difference multiple comparisons for group phase by friendship relation.

Group Type I	Group Type J	Mean		p-value	Lower Bound	Upper Bound
		Difference (I - J)	Std. Error			
No relationship	Enemies	0.083	0.139	0.934	-0.276	0.442
	Friendship one-way	0.066	0.124	0.950	-0.253	0.386
	Both friends	-0.251	0.117	0.138	-0.552	0.049
Enemies	Friendship one-way	0.149	0.150	0.751	-0.236	0.535
	Both friends	-0.168	0.144	0.645	-0.538	0.202
Friendship one-way	Both friends	-0.318*	0.129	0.066	-0.650	-0.014

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the Tukey's HSD test results for the mean differences, standard deviations, p-values, and 95% confidence intervals across all friendship relations in the group phase for low-ability students.

Table A2.9: T-tests for mean differences in scores by gender.

Test	Mean Difference	Std. Error	p-value	Lower Bound	Upper Bound
Baseline Score	0.3675***	0.0786	<0.0001	0.2133	0.5217
Group Score	0.1091	0.0714	0.1266	-0.0309	0.2490
Final Score	0.3977***	0.1017	0.0001	0.1982	0.5973

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the t-test results for the mean differences, standard deviations, p-values, and 95% confidence intervals for math scores between male and female groups.

Table A2.10: Tukey's Honestly Significant Difference multiple comparisons for final test by gender

Group Type I	Group Type J	Mean		p - value	Lower Bound	Upper Bound
		Difference (I - J)	Std. Error			
Both male	Female, partner male	0.200	0.144	0.505	-0.170	0.570
	Male, partner female	-0.149	0.144	0.730	-0.520	0.222
	Both female	0.442**	0.141	0.009	0.080	0.804
Both female	Female, partner male	-0.242	0.143	0.330	-0.611	0.127
	Male, partner female	-0.349	0.147	0.082	-0.727	0.029
Female, partner male	Male, partner female	0.349	0.147	0.082	-0.029	0.727

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the Tukey's HSD test results for the mean differences, standard deviations, p-values, and 95% confidence intervals across all gender pairings in the final phase test.

Table A2.11: T-tests for mean differences in scores by bullying group

Test	Mean Difference	Std. Error	p-value	Lower Bound	Upper Bound
Baseline Score	0.4504***	0.1107	<0.0001	0.2334	0.6675
Group Score	0.3590***	0.1001	0.0003	0.1628	0.5553
Final Score	-0.0597	0.1433	0.6768	-0.3408	0.2213

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the t-test results for the mean differences, standard deviations, p-values, and 95% confidence intervals for math scores between students who do and do not experience bullying.

Table A2.12: ANOVA results for all test phases by bullying status

Score Type	Observations	Model SS	Residual SS	Total SS	F-value	p-value
Baseline Score	1,640	42.142***	4166.467	4208.609	16.57	0.000
Group Score	1,638	47.862***	3367.985	3415.847	11.62	0.000
Final Score	1,598	20.291	6640.104	6660.396	1.62	0.182

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the ANOVA results for all bullying pairings in the three test phases.

Table A2.13: T-tests for mean differences in math scores by migrant status

Test	Mean Difference	Std. Error	p-value	Lower Bound	Upper Bound
Baseline Score	0.1079	0.1589	0.4968	-0.2036	0.4196
Group Score	0.0811	0.1439	0.5730	-0.2011	0.3632
Final Score	0.8789***	0.2051	<0.0001	0.4767	1.2812

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the t-test results for mean differences, standard deviations, p-values, and 95% confidence intervals for math scores between local and migrant students.

Table A2.14: ANOVA Results for Scores by migrant status and partner's migrant status

Score Type	Observations	Model SS	Residual SS	Total SS	F-value	p-value
Baseline Score Math	1,640	1.187	4,207.423	4,208.609	0.46	0.497
Group Score Math	1,638	1.050	3,414.797	3,415.847	0.25	0.778
Final Score Math	1,598	81.063***	6,579.332	6,660.396	6.55	0.0002

This table presents the ANOVA results for all migrant pairings in the three test phases.

Table A2.15: Tukey's Honestly Significant Difference multiple comparisons Test for final math scores by migrant status

Group Type I	Group Type J	Mean Difference (I - J)	Std. Error	p-value	Lower Bound	Upper Bound
Migrant, partner local	Both local	-0.850***	0.2233	0.001	-1.4241	-0.2758
Local, partner migrant	Both local	-0.228	0.2233	0.738	-0.8019	0.3464
	Migrant, partner local	0.622	0.3063	0.177	-0.1655	1.4098
Both migrant	Both local	-1.112	0.4957	0.112	-2.3870	0.1627
	Migrant, partner local	-0.262	0.5382	0.962	-1.6465	1.1219
	Local, partner migrant	-0.884	0.5382	0.355	-2.2686	0.4998

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the Tukey's HSD test results for the mean differences, standard deviations, p-values, and 95% confidence intervals across all migration status pairings in the final phase test.

Table A2.16: ANOVA Results for Comfort Level by Quartile and Migrant Status

Factor	Observations	Model SS	Residual SS	Total SS	F-value	p-value
Quartile	1,628	1.561	539.323	540.885	1.57	0.195
Migrant Status	1,628	0.494	540.391	540.885	0.49	0.686

This table presents the ANOVA results for comfort depending on students quartile and migration status. The results indicate no statistically significant differences at conventional significance levels in comfort levels across these categories.

Table A2.17: Tukey's Honestly Significant Difference multiple comparisons for comfort level by friendship relation

Group Type I	Group Type J	Mean Difference (I - J)	Std. Error	p - value	Lower Bound	Upper Bound
No relationship	Enemy	0.2685***	0.0408	0.000	0.1727	0.3643
Friend	Enemy	0.4217***	0.0378	0.000	0.3330	0.5104
	No relationship	0.1532***	0.0313	0.000	0.0797	0.2268

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the Tukey's HSD test results for the mean differences, standard deviations, p-values, and 95% confidence intervals across friendship pairings for the comfort level.

Table A2.18: Tukey's Honestly Significant Difference multiple comparisons for comfort level by group preferences

Group Type I	Group Type J	Mean		<i>p</i> - value	Lower Bound	Upper Bound
		Difference (I - J)	Std. Error			
Does not want to work with partner	No preferences	0.3111***	0.0348	0.000	0.2295	0.3926
	Wants to work with partner	0.3542***	0.0394	0.000	0.2618	0.4466
No preferences	Wants to work with partner	0.0431	0.0333	0.399	-0.0351	0.1213

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the Tukey's HSD test results for the mean differences, standard deviations, *p*-values, and 95% confidence intervals across group preference pairings for the comfort level.

Table A2.19: ANOVA Results for Math Scores by wealth level

Score Type	Observations	Model SS	Residual SS	Total SS	<i>F</i> -value	<i>p</i> -value
Baseline Score Math	1,627	1.036	4,168.608	4,169.644	0.20	0.817
Group Score Math	1,612	5.076	3,349.974	3,355.050	1.22	0.296
Final Score Math	1,572*	57.592	6,511.102	6,568.694	1.73	0.087

This table presents the ANOVA results for math scores across different wealth groups in each test phase. Final test shows a weakly statistical difference across some groups, so we perform a Tukey's HSD test for this specific test score.

Table A2.20: Tukey's Honestly Significant Difference multiple comparisons test for final math scores by wealth subgroup

Group Type I	Group Type J	Mean		<i>p</i> -value	Lower Bound	Upper Bound
		Difference (I - J)	Std. Error			
Both below average	Below average, partner above average	0.243	0.359	0.999	-0.872	1.358
	Below average, partner average	0.274	0.433	0.999	-1.071	1.619
	Average, partner below average	0.460	0.359	0.936	-0.654	1.574
	Both average	0.641	0.332	0.593	-0.391	1.672
	Average, partner above average	0.733	0.357	0.507	-0.376	1.842
	Above average, partner below average	0.713	0.429	0.770	-0.620	2.047

Continued on next page

Table A2.20: (continued)

Group Type I	Group Type J	Mean Difference (I - J)	Std. Error	p-value	Lower Bound	Upper Bound
	Above average, partner average	0.729	0.357	0.515	-0.381	1.839
	Both above average	0.918	0.424	0.429	-0.399	2.236
Below average, partner above average	Below average, partner average	0.031	0.329	1.000	-0.989	1.052
	Average, partner below average	0.217	0.222	0.988	-0.471	0.906
	Both average	0.398	0.175	0.362	-0.147	0.943
	Average, partner above average	0.490	0.219	0.383	-0.191	1.171
	Above average, partner below average	0.470	0.324	0.877	-0.535	1.476
	Above average, partner average	0.486	0.220	0.396	-0.196	1.168
	Both above average	0.675	0.317	0.452	-0.309	1.659
Below average, partner average	Average, partner below average	0.186	0.328	1.000	-0.834	1.206
	Both average	0.367	0.299	0.951	-0.562	1.296
	Average, partner above average	0.459	0.327	0.896	-0.556	1.473
	Above average, partner below average	0.439	0.404	0.976	-0.816	1.695
	Above average, partner average	0.455	0.327	0.901	-0.560	1.470
	Both above average	0.644	0.399	0.797	-0.595	1.883
Both average	Average, partner above average	0.181	0.175	0.983	-0.363	0.724
	Average, partner above average	0.273	0.219	0.946	-0.407	0.952
	Above average, partner below average	0.253	0.323	0.997	-0.752	1.258
	Above average, partner average	0.269	0.219	0.951	-0.412	0.950
	Both above average	0.458	0.317	0.880	-0.526	1.442
	Both average	0.092	0.172	1.000	-0.442	0.626
Above average, partner average					Continued on next page	

Table A2.20: (continued)

Group Type I	Group Type J	Mean		<i>p</i> -value	Lower Bound	Upper Bound
		Difference (I - J)	Std. Error			
	Above average, partner below average	0.073	0.294	1.000	-0.840	0.985
	Above average, partner average	0.088	0.172	1.000	-0.447	0.623
	Both above average	0.277	0.286	0.989	-0.611	1.166
Above average, partner below average	Above average, partner average	-0.020	0.322	1.000	-1.019	0.980
	Both above average	0.185	0.315	1.000	-0.793	1.163
	Above average, partner average	0.016	0.322	1.000	-0.984	1.016
	Both above average	0.205	0.395	1.000	-1.021	1.431

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the Tukey's HSD test results for the mean differences, standard deviations, *p*-values, and 95% confidence intervals across all wealth group pairings for the final math test scores, with no significant differences across any groups.

Table A2.21: T-tests for sample mean across group tests

Test	Mean Difference	Std. Error	<i>p</i> -value	Lower Bound	Upper Bound
Baseline Score	0.0.013358	0.097	0.7339	-0.1563	0.2218
Group Score	-0.0.013	0.085	0.8788	-0.1761	0.1507
Final Score	0.0.014	0.127	0.9122	-0.2244	0.2511

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the t-test results for mean differences, standard deviations, *p*-values, and 95% confidence intervals for math scores between students and their partners, with none being statistically significant.

Table A2.22: Levene's Test for Homogeneity of Variance

Test Phase	Df	F value	<i>p</i> -value	N
Baseline Score	1	2.6519	0.1037	1205
Group Score	1	0.0324	0.8573	1205
Final Score	1	0.3032	0.5820	1181

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the results from Levene's test for homogeneity of variance for math scores between students and their partners, with none being statistically significant.

Table A2.23: Summary statistics for individuals suffering from bullying

Variable	Mean	Std. Dev.	Min	Max
Age	13.92	2.81	11	33
Bullying Popularity	1.71	1.40	1	9
Self-Bullying	0.16	0.37	0	1
Quartile	2.13	1.04	1	4
Migrant	0.04	0.21	0	1
Female	0.40	0.49	0	1
N	247			

This table presents the main summary statistics for the students suffering bullying.

Table A2.24: Distribution of Friends and Group vs. Enemies and No Group

Category	Number of coincidences	Frequency	Percentage	Cumulative %
Friends and Group	0	27	1.61	1.61
	1	29	1.73	3.34
	2	78	4.65	8.00
	3	128	7.64	15.63
	4	251	14.98	30.61
	5	1,163	69.39	100.00
Enemies and No Group	0	150	8.95	8.95
	1	185	11.04	19.99
	2	275	16.41	36.40
	3	298	17.78	54.18
	4	292	17.42	71.60
	5	476	28.40	100.00

This table presents the number of students that were selected both as a friend and as someone you want in your group (with a maximum of 5) and those selected both as an enemy and someone you do not want in your group (with a maximum of 5).

Table A2.25: Regression of individual characteristics and relationship types on group performance

	(1)	(2)	(3)
Number of friends	0.00299 (0.00407)	0.00417 (0.00397)	0.00430 (0.00356)
Number of times selected as friend	0.0117 (0.00809)	0.0105 (0.00782)	0.00797 (0.00695)
Score baseline test		0.196*** (0.0209)	0.192*** (0.0264)
Score baseline test partner		0.176*** (0.0209)	0.168*** (0.0275)
Friendship both ways		-0.237*** (0.0772)	-0.254** (0.105)
Friendship one-way		-0.168** (0.0740)	-0.162 (0.110)
Enemy with friend		0.117 (0.212)	0.102 (0.320)
Enemy one-way		-0.0699 (0.108)	-0.0876 (0.159)
Enemy both ways		-0.417** (0.173)	-0.413* (0.211)
Class FE	✓	✓	✓
Controls	×	×	✓
Observations	1635	1635	1609
Mean of Dep. Variable	3.645	3.645	3.650

Standard errors in parentheses, clustered by class.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Omitted category is “No relationship between students”

Table A2.26: Summary Statistics for students

Variable	High-ability students			Low-ability students		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Group Test						
Number of Friends	15.02	9.23	279	13.54	9.61	272
Selected as Friend	12.52	5.43	279	11.11	5.99	272
Female	0.43	0.50	279	0.57	0.50	272
Migrant	0.08	0.26	279	0.09	0.29	272
Bullying Popularity	0.15	0.64	279	0.30	0.90	272
Final Test						
Number of Friends	15.66	9.99	289	11.90	7.69	268
Selected as Friend	12.43	5.34	289	10.22	5.28	268
Female	0.42	0.49	289	0.59	0.49	268
Migrant	0.06	0.24	289	0.09	0.29	268
Bullying Popularity	0.21	0.76	289	0.29	0.90	268

This table presents the main summary statistics for low-ability students who improved their performance during the group/final test with respect to their score in the baseline test and high-ability students who decreased their performance during the group/final test with respect to their score in the baseline test.

Table A2.27: T-tests for sample mean across group tests

	Mean Difference	Std. Error	p-value	Lower Bound	Upper Bound
Low-ability improving in group phase					
Number of Friends	-2.025***	0.706	0.0042	-3.411	-0.640
Selected as Friend	-1.127***	0.411	0.0062	-1.932	-0.321
Female	0.047	0.036	0.1936	-0.024	0.118
Migrant	0.037**	0.018	0.0421	0.001	0.072
Bullying Popularity	-0.012	0.063	0.8539	-0.136	0.112
Low-ability improving in final phase					
Number of Friends	0.283	0.706	0.6888	-1.103	1.668
Selected as Friend	0.128	0.410	0.7551	-0.677	0.934
Female	0.019	0.036	0.5921	-0.051	0.090
Migrant	0.035*	0.018	0.0516	-0.000	0.071
Bullying Popularity	-0.002	0.063	0.9720	-0.126	0.121
High-ability decreasing in group phase					
Number of Friends	1.214*	0.698	0.0824	-0.156	2.584

Continued on next page

Table A2.27: (continued)

	Mean Difference	<i>Std. Error</i>	<i>p-value</i>	Lower Bound	Upper Bound
Selected as Friend	1.038**	0.410	0.0116	0.233	1.843
Female	-0.009	0.038	0.8140	-0.084	0.066
Migrant	0.015	0.019	0.4205	-0.022	0.053
Bullying Popularity	-0.079	0.057	0.1651	-0.190	0.033
High-ability decreasing in final phase					
Number of Friends	2.329***	0.690	0.0008	0.975	3.684
Selected as Friend	0.910**	0.408	0.0260	0.109	1.711
Female	-0.028	0.038	0.4566	-0.102	0.046
Migrant	-0.006	0.019	0.7472	-0.043	0.031
Bullying Popularity	0.033	0.056	0.5600	-0.078	0.144

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the t-test results for mean differences, standard deviations, p-values, and 95% confidence intervals for various metrics across different groups. Specifically, we compare low-ability students improving in a given phase with the rest of low-ability students not improving in that phase, as well as high-ability students decreasing their performance in a given phase with the rest of high-ability students that either maintained or increased their performance in that phase.

Table A2.28: Tukey's Honestly Significant Difference multiple comparisons for performance by group composition

Group Type I	Group Type J	Mean Difference (I - J)	Std. Error	p - value	Lower Bound	Upper Bound
Low-Ability Male Students						
2 Male low	Male low - Female low	0.2238	0.1405	0.383	-0.1380	0.5857
	Male high - Male low	0.7536***	0.1902	0.000	0.2635	1.2437
Male low - Female low	Male high - Male low	0.5298**	0.1943	0.033	0.0291	1.0304
	Female high - Male low	0.0203	0.1854	1.000	-0.4574	0.4980
Female high - Male low	Male high - Male low	-0.5094	0.2255	0.109	-1.0903	0.0715
High-Ability Male Students						
2 Male high	Male high - Female high	-0.3701	0.2917	0.584	-1.1253	0.3851
	Male high - Male low	-0.5677	0.2631	0.138	-1.2487	0.1134
	Male high - Female low	-0.5566	0.2697	0.168	-1.2547	0.1415
Male high - Female high	Male high - Male low	-0.1976	0.2491	0.857	-0.8424	0.4472
	Male high - Female low	-0.1865	0.2560	0.886	-0.8492	0.4762
Male high - Female low	Male high - Male low	0.0111	0.2228	1.000	-0.5657	0.5878

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the Tukey's HSD test results for the mean differences, standard deviations, p-values, and 95% confidence intervals across group compositions for performance.

Table A2.29: Tukey's Honestly Significant Difference multiple comparisons for performance by group composition

Group Type I	Group Type J	Mean Difference (I - J)	Std. Error	p - value	Lower Bound	Upper Bound
Low-Ability Female Students						
2 Female low	Male low - Female low	0.2875	0.1311	0.126	-0.0502	0.6252
	Female high - Female low	0.2935	0.1852	0.388	-0.1837	0.7706
	Male high - Female low	0.6775***	0.1906	0.002	0.1863	1.1687
Male low - Female low	Female high - Female low	0.0060	0.1901	1.000	-0.4838	0.4958
	Male high - Female low	0.3900	0.1954	0.191	-0.1135	0.8935
Female high - Female low	Male high - Female low	0.3840	0.2352	0.361	-0.2219	0.9899
High-Ability Female Students						
2 Female high	Male high - Female high	0.0766	0.3141	0.995	-0.7359	0.8891
	Female high - Female low	-0.5543	0.2843	0.210	-1.2898	0.1811
	Female high - Male low	-0.4817	0.2731	0.293	-1.1881	0.2247
Male high - Female high	Female high - Female low	-0.6310	0.2862	0.125	-1.3713	0.1094
	Female high - Male low	-0.5583	0.2751	0.180	-1.2698	0.1532
Female high - Male low	Female high - Female low	0.0727	0.2405	0.990	-0.5495	0.6948

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents the Tukey's HSD test results for the mean differences, standard deviations, p-values, and 95% confidence intervals across group compositions for performance.

4 Employment Protection & Education Decisions

4.1 Introduction

The relationship between dismissal costs and labor market outcomes has long attracted attention from economists and policymakers. Although early theoretical models and empirical studies have shown that higher firing costs can reduce employment in certain sectors, subsequent research has revealed that the impact of these policies is highly context-dependent. Some industries, especially those relying on specialized human capital, may even benefit from increased employment protection. Yet, while these studies have extensively examined the direct effects of firing costs on employment and wages, much less is known about the mechanisms through which these policies might shape broader labor market dynamics, including the educational choices that individuals make.

Education plays a dual role in modern labor markets. On the one hand, it directly enhances an individual's productivity by developing valuable skills and knowledge; on the other hand, it functions as a signal to employers, conveying information about unobservable abilities. In a competitive hiring environment, the relative importance of these functions may shift in response to regulatory changes. Given that higher termination costs make it more challenging to dismiss unsatisfactory employees, employers are likely to rely more heavily on educational credentials as a pre-hiring signal of worker quality. In other words, because it becomes costlier to terminate a poorly performing employee, firms place significant emphasis on an applicant's education during the hiring process to mitigate the risk of future costly dismissals. This, in turn, might alter both the returns to education—as reflected in wage differentials—and the incentives for individuals to pursue additional schooling.

Despite the considerable theoretical interest in these issues, empirical evidence on the impact of dismissal costs on education decisions remains scarce. It is not clear whether changes in employment protection trigger a reallocation of educational investments, or whether any observed adjustments in the returns to education are simply a byproduct of shifts in labor market behavior. In other words, do higher firing costs encourage workers to invest in more education as a compensatory signal, or do they primarily influence labor outcomes through mechanisms unrelated to educational attainment? Answering these questions is critical for understanding the broader implications of labor market reforms on human capital formation and economic inequality.

Other studies have documented that dismissal cost policies have heterogeneous impacts on labor market outcomes. For instance, [Autor et al. \(2006\)](#) and [Macleod and Nakavachara \(2007\)](#) provide evidence that firing cost reforms can reduce employment rates in certain sectors while potentially boosting job security and on-the-job training in others. Their work, however, has primarily focused on the direct wage and employment effects, leaving the influence on educational decisions largely unexamined. This gap is significant because the educational attainment process itself may adjust to the changing incentives created by increased dismissal costs.

In a related vein, a considerable body of work examines the dual role of education as both an investment in human capital and a signal of underlying ability. [Spence \(1973\)](#) highlighted that educational credentials serve as a proxy for productivity in environments characterized by information asymmetry. Despite the robust theoretical foundation, empirical research integrating these two functions with labor market policy adjustments has been relatively scarce. Bridging this gap is essential for understanding whether higher dismissal costs intensify the signaling function of education and, consequently, reshape the incentives for pursuing additional schooling.

Furthermore, research into the microfoundations of labor market behavior suggests that workers adjust their educational strategies in response to shifts in the economic environment. For example, [Becker \(1964\)](#) and [Mincer \(1974\)](#) laid the theoretical groundwork for understanding human capital investment, arguing that individuals weigh expected wage premiums against the costs of schooling. Empirical studies such as [Card \(1999\)](#) and [Oreopoulos \(2006\)](#) provide evidence that fluctuations in wage structures lead workers to recalibrate their education investments, with individuals more likely to pursue additional schooling when the anticipated returns are higher. However, while these studies document adjustments in educational investment in response to changes in expected returns, most analyses following employment protection reforms (like those by [Autor et al. \(2006\)](#) and [Macleod and Nakavachara \(2007\)](#)) have concentrated on wage differentials rather than directly linking these shifts to systematic changes in educational attainment. This gap is particularly salient given the persistent concerns about overeducation and skill mismatches in modern labor markets, as highlighted by [Brynin \(2002\)](#).

Finally, several studies have examined the broader institutional context in which these policies operate. [Cabrales et al. \(2017\)](#) and related works have underscored that employment protection may indirectly influence factors such as on-the-job training and career mobility. Yet, few have extended this analysis to consider the interplay between dismissal costs and education decisions. By integrating insights from both labor market and educational signaling literatures, our study contributes a novel perspective: it not only evaluates the direct labor market effects of firing cost regulations but also uncovers how these policies may lead workers to alter their educational trajectories. This integrated approach offers a more comprehensive understanding of the welfare implications of employment protection reforms. Our study directly connects dismissal cost policies to both the wage premium for education and the educational choices made by workers. In doing so, we clarify how changes in employer behavior—prompted by policy-induced changes—can shift incentives for investing in human capital, ultimately influencing long-term economic outcomes and individual career paths.

In this paper, we analyze the changes in U.S. employment protection policies that began in the 1970s, a pivotal shift away from the traditional employment-at-will doctrine. Historically, employers were free to terminate employees without cause, but increasing concerns over job security led state courts to impose limits on arbitrary dismissals. These legal reforms are embodied in three principal doctrines: the implied contract exception, which holds that certain employer assurances or customary practices create an expectation of job stability; the public policy exception, which prevents terminations that contravene essential societal interests; and the implied covenant of good faith and fair dealing, which ensures that dismissals are not executed in a manner that violates the spirit of fair treatment.

[Dertouzos and Karoly \(1993\)](#) provide early evidence that the introduction of these employment protection doctrines was associated with a significant decline in employment rates, highlighting the broad impact of these policies on labor market dynamics. By scrutinizing these legal changes, our study not only evaluates the direct effects on employment outcomes but also investigates how such policies may influence the returns to education and the educational choices made by workers.

Our results indicate that the implementation of dismissal cost laws exerts a negative effect on the returns to education, with this effect diminishing as age increases. In our analysis, the presence of any employment protection appears to lower the premium associated with educational attainment, particularly among younger cohorts who lack the buffering influence of accumulated work experience.

When we disaggregate the different types of dismissal cost laws, the implied contract exception emerges as producing the most pronounced effect among the three doctrines considered. However, some of the distinct effects associated with this exception tends to dissipate when controlling for the overall presence of a protection law. This pattern suggests that it is the existence of some form of dismissal protection that drives the observed reduction in education returns, rather than any one particular doctrine on its own. The presence of this exception also presents a positive effect, albeit very small, on the probability of being employed, suggesting the presence of this exception reduces the number of dismissals.

In contrast, our investigation into the effect of these policies on educational attainment decisions reveals no significant direct impact. The differences in education levels between the treatment and control groups are minimal and statistically insignificant, reinforcing the notion that the changes in the returns to education are not accompanied by corresponding adjustments in individuals' educational choices. This finding implies that, while the labor market rewards for education are altered by the adoption of dismissal cost laws, individuals do not modify their education investments in response.

Overall, these results suggest that the labor market effects of dismissal cost policies are primarily a direct consequence of the regulatory environment on wages and employment outcomes, rather than an indirect effect mediated by changes in educational attainment. The robustness of these findings across alternative specifications further supports the conclusion that, while the returns to education are sensitive to changes in employment protection, the decision-making process regarding educational investment remains largely unaffected.

The remainder of the paper is organized as follows. In [Section 2](#), we outline the theoretical framework that underpins our analysis, detailing the mechanisms by which dismissal cost laws may affect both labor market outcomes and educational returns. [Section 3](#) provides the institutional and policy background, describing the evolution of employment protection and the specific legal doctrines that form the basis of our study. [Section 4](#) explains the data sources and construction of our empirical dataset, while [Section 5](#) details the methodological approach employed to estimate the effects of dismissal cost laws. In [Section 6](#), we present our main empirical findings along with a series of robustness checks that validate our results. Finally, [Section 7](#) summarizes and concludes.

4.2 Theoretical framework

In recent decades, the average educational attainment of the workforce has risen substantially, as highlighted by [Snyder \(1993\)](#) and [Ortiz-Ospina and Roser \(2016\)](#). As a consequence, a growing share of employees find themselves performing tasks for which their qualifications exceed job requirements, as shown in [Brynin \(2002\)](#) and [Arvan et al. \(2019\)](#). Such overeducation entails social costs, whether borne by individuals or the public sector, like potential job dissatisfaction and productivity losses, and it is aggravated even more if we consider the cost of opportunity of this overeducation.

Spence (1973)'s signalling model offers a theoretical lens through which to interpret this phenomenon. In this framework, additional education may or may not raise an individual's productivity; instead, it functions as a costly signal of unobserved ability. By obtaining higher education levels, candidates differentiate themselves from lower-ability applicants, enabling employers, who cannot directly observe worker ability, to screen more effectively.

Returns to education can thus be decomposed into two distinct components. Under the human-capital paradigm, schooling directly augments an individual's productive capacity, generating greater output and higher wages. By contrast, under the signalling paradigm, education yields monetary rewards primarily because employers use educational attainment to screen job applicants when they cannot directly observe ability. Importantly, the strength of the signalling mechanism depends on labor-market institutions: when hiring becomes more rigid (due to stricter employment-protection legislation, shorter probationary periods or high firing costs) firms place greater weight on educational credentials to minimize the risk of hiring a low-ability employee.

Our empirical framework builds on the model of Saint-Paul (1997), which distinguishes two worker types, low-skilled and high-skilled, whose true skill levels are unobservable to firms *ex ante*. Firms always prefer to employ high-skilled workers and, when firing costs are low, will readily replace a low-skilled incumbent with a high-skilled applicant. Saint-Paul shows that a rise in firing costs has two countervailing effects on returns to education. First, when replacement is difficult, firms may respond by investing in on-the-job training for incumbent low-skilled workers, increasing their human-capital accumulation and thus the measured return to education; however, this can also lead to a larger skilled workforce and, paradoxically, a higher equilibrium unemployment rate. Second, if higher firing costs induce workers to acquire more formal education prior to hiring, therefore raising overall productivity, the economy may reach a new equilibrium with both higher returns to education and lower unemployment.¹

However, how do changes in firing costs translate into altered returns to education and individuals' decisions regarding their level of education? This paper addresses these questions. If employment protection directly affects returns to education, we should observe two patterns around a policy reform: an immediate shift in educational returns (driven by signalling) and a subsequent change in schooling attainment (driven by human-capital accumulation). Kugler and Saint-Paul (2004) show that, following an increase in firing costs, firms prefer to recruit already-employed workers, who signal a lower probability of being low-skilled workers, over the unemployed, consistent with the theoretical model. Accordingly, we hypothesize two main effects:

1. A contemporaneous effect on workers at the time a state raises firing costs. As firms worry more about distinguishing high-skilled from low-skilled workers, the signalling value of additional education rises, boosting its returns.
2. An effect on young individuals still in education when firing costs increase. Anticipating stronger signalling requirements in the labor market, these cohorts should respond by extending their educational attainment in the years following the reform.

¹For a deeper explanation of the model, we refer the reader to Chapters 9 and 10 of Saint-Paul (1997).

4.3 Institutional framework

4.3.1 Employment at will

Employment in the U.S. was governed by the employment-at-will doctrine, first articulated in 1884 by the Tennessee Supreme Court, which held that “men must be left, without interference, to buy and sell where they please, and to discharge or retain employees at will for good cause or for no cause, or even for bad cause without thereby being guilty of an unlawful act *per se*.”² Over the course of the 20th century, this principle was progressively constrained, as can be seen in Figure 4.1. California led the shift in 1959 by recognizing a judicial exception to at-will employment. Henceforth, during the 1970s and 1980s, forty-four additional states adopted similar exceptions; by 1992, forty-six of the fifty states had modified the doctrine, the majority doing so between 1980 and 1985.

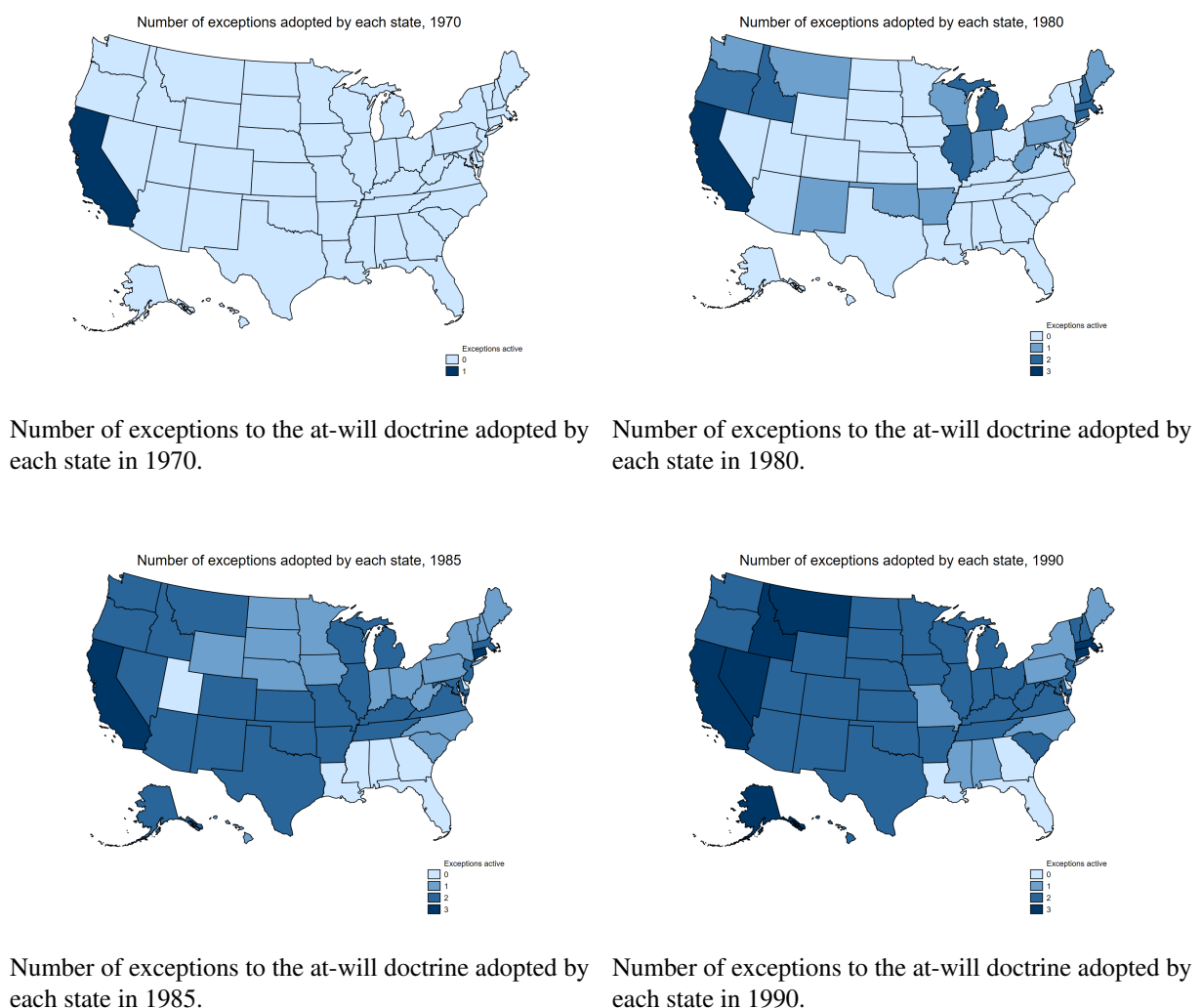


Fig. 4.1. Adoption of the at-will doctrines adopted by each state in the U.S. across time.
Source: own elaboration using the data from Autor et al. (2006)

²Payne v. Western & Atlantic Railroad, Tennessee 1884.

4.3.2 Exceptions of the employment at will doctrine

By the end of the 20th century, three exceptions to the at-will doctrine had become widely recognized³:

1. **Implied Contract Exception:** A court will infer a legally binding employment contract when an employer's conduct or communications create a reasonable expectation of job security. Examples include employee handbooks promising termination only for just cause, consistent patterns of salary increases and promotions, or customary industry practices indicating long-term retention.
2. **Public Policy Exception:** Employers may not terminate an employee for refusing to abide by fundamental public policies, such as serving on a jury, refusing to commit perjury, etc.
3. **The Implied Covenant of Good Faith and Fair Dealing:** An employer may not discharge an employee for "bad cause", such as depriving the employee of earned benefits, for example, firing just before a sales bonus or commission is due.

4.4 Data

Our primary data sources are the U.S. Census and the Current Population Survey (CPS). The Census offers a detailed, nationally representative cross-section of the U.S. population from 1950 to 2018, comprising: a 1% sample in 1950 and 1970; a 5% sample in 1960, 1980, 1990 and 2000; and, from 2001 onward, the American Community Survey (ACS), which maintains a 1% annual sample. For each respondent we observe state of birth, state of residence, age, current student status, highest educational attainment, and a range of socioeconomic characteristics (gender, race, marital status, employment status, earnings, poverty status, etc.).

To estimate the impact of firing-cost reforms on returns to education, we supplement the Census with the March Annual Social and Economic Supplement (ASEC) over the period 1977–2009. The ASEC provides, in addition to basic demographic and educational measures, richer labor-market variables for a subsample of respondents, such as weeks unemployed in the prior year, employment status (full- versus part-time), and hours worked in the reference week. The first year we use is 1977 because that is the first year the CPS reports respondents' state of residence (earlier years identify only broader "areas"). Given that our latest firing-cost change occurs in 1999, we adopt a ten-year window around each reform.

All datasets are merged with the legal appendix from [Autor et al. \(2006\)](#), which precisely records the month in which each state adopted its wrongful-discharge law and the prevalence of each exception⁴. From these dates, we construct indicator variables for the presence of each legal exception in the respondent's state at the time of interview.

In our analysis of returns to education, focusing on labor-market outcomes such as employment rates and wages, we assign treatment based on the respondent's state of residence, as this is the jurisdiction in which they are most likely employed. By contrast, when examining educational-attainment decisions, we use the state of birth to determine exposure to the policy reform: since roughly 30% of individuals reside in a different state from their birth, and migrations typically occur after educational choices are

³Table [Table A4.8](#) in the appendix reports, for each state in our sample, the year an exception was adopted (and, where applicable, period of effect), as well as the number of pre- and post-treatment years available.

⁴For more information about this, we refer the reader to [Autor et al. \(2006\)](#).

made, the timing of doctrine adoption in the birth state is more likely to better capture the causal effect on schooling decisions. Importantly, results using state of residence instead of birth are similar, underscoring the robustness of our estimates. For all robustness checks involving both the Census and CPS samples, we rely on state of residence, since the state of birth is not available in the CPS.

The Census sample initially comprised 100,663,954 observations. We first excluded records for individuals born or residing outside U.S. states with documented doctrine-adoption dates, as well as respondents still of studying age at the time of interview (and thus without a final attainment level). This screening reduced the sample to 48,650,284 observations. To improve computational efficiency, we then collapsed the data into cells defined by survey year, state of birth, birth cohort, sex and race, taking the cell-mean of all relevant outcomes. Next, for each treated state we created a “group indicator” that includes both the treated state and its corresponding set of control states. This step requires duplicating some cells across multiple groups. The resulting Census-based dataset contains 477,327 observations.

The CPS sample began with 9,753,610 person-year observations. Applying analogous exclusions (dropping states without firing-cost data, respondents younger than 16 or older than 40 at the moment of the interview, and all observations after 2009 to maintain a consistent treatment window) yields a working sample of 1,838,654 observations.

Data description for selected variables

Variable	Description
Education	Maximum level of education attained, as reported by the respondent
High School	Dummy which takes value 1 if the maximum level of education attained by the individual is no more than High School graduate
College	Dummy which takes value 1 if the maximum level of education attained by the individual is at least some year of College
Treated State	Dummy which takes value 1 if the individual was born/lives at the time of the interview in a state in the treatment group
Age	Age at the moment of interview, as reported by the respondent
Age X	Dummy which takes value 1 if the individual belongs to age group X (15-19 years old, 20-24 years old, etc.), at the time of treatment
Female	Dummy which takes value 1 if the individual is a female
Black/Asian/Other race	Dummy which takes value 1 if the individual belongs to ethnicity X
Wage	Individual's self-reported total pre-tax wage and salary income for the previous calendar year, in dollars of 1999
Total Income	Individual's self-reported total pre-tax personal income or losses for the previous calendar year, in dollars of 1999
G/C/P	Dummy which takes value one if policy G/C/P (affecting firing costs) is implemented in the individual's state at the time of interest
Employed	Dummy which takes value 1 if the individual was employed for at least some week during the previous calendar year

Descriptive statistics for Census data (means by groups)					Descriptive statistics for CPS data				
	Mean	SD	Min	Max		Mean	SD	Min	Max
Years of education	8.51	1.31	0	12.00	Years of education	7.62	2.31	0	12.00
High School	0.51	0.25	0	1.00	Wage (in thousand \$)	17.7	23.5	0	629.9
College	0.39	0.23	0	1.00	Employed	0.80	0.40	0	1.00
Treated State	0.30	0.46	0	1.00	High School	0.57	0.49	0	1.00
Age	51.20	10.28	31	81.00	College	0.43	0.49	0	1.00
Age 15-19	0.18	0.39	0	1.00	Age	28.00	7.29	16	40.00
Age 20-24	0.19	0.40	0	1.00	Age 15-19	0.17	0.38	0	1.00
Age 25-29	0.20	0.40	0	1.00	Age 20-24	0.19	0.39	0	1.00
Age 30-40	0.43	0.49	0	1.00	Age 25-29	0.20	0.40	0	1.00
Female	0.50	0.50	0	1.00	Age 30-40	0.45	0.50	0	1.00
Black	0.29	0.45	0	1.00	Female	0.52	0.50	0	1.00
Asian	0.12	0.33	0	1.00	Black	0.11	0.31	0	1.00
Other ethnic group	0.29	0.45	0	1.00	Asian	0.02	0.15	0	1.00
					Other ethnic group	0.03	0.17	0	1.00

Census (means by groups) by age group					CPS by age group				
	15-19	20-24	24-29	30-40		16-19	20-24	25-29	30-40
Years of education	8.72 (1.32)	8.60 (1.28)	8.54 (1.26)	8.36 (1.32)	Years of education	5.68 (1.39)	7.60 (1.85)	8.06 (2.27)	8.18 (2.39)
High School	0.47 (0.25)	0.50 (0.25)	0.51 (0.24)	0.53 (0.25)	Wage (in thousand \$)	2.6 (5.6)	11.3 (12.6)	19.6 (19.4)	25.3 (28.9)
College	0.42 (0.24)	0.40 (0.23)	0.39 (0.23)	0.36 (0.23)	Employed	0.58 (0.49)	0.82 (0.38)	0.84 (0.36)	0.85 (0.36)
Age	42.53 (6.48)	46.33 (7.40)	50.57 (8.11)	57.38 (9.64)	High School	0.91 (0.29)	0.53 (0.50)	0.51 (0.50)	0.49 (0.50)
Female	0.50 (0.50)	0.50 (0.50)	0.50 (0.50)	0.50 (0.50)	College	0.09 (0.29)	0.47 (0.50)	0.49 (0.50)	0.51 (0.50)
Black	0.27 (0.45)	0.28 (0.45)	0.29 (0.45)	0.29 (0.46)	Age	17.42 (1.10)	22.01 (1.42)	27.02 (1.41)	34.97 (3.16)
Asian	0.16 (0.36)	0.14 (0.35)	0.12 (0.33)	0.10 (0.29)	Female	0.49 (0.50)	0.52 (0.50)	0.52 (0.50)	0.52 (0.50)
Other ethnic group	0.28 (0.45)	0.28 (0.45)	0.29 (0.45)	0.29 (0.45)	Black	0.12 (0.33)	0.11 (0.32)	0.10 (0.31)	0.10 (0.29)
					Asian	0.02 (0.13)	0.02 (0.14)	0.02 (0.15)	0.03 (0.16)
					Other ethnic group	0.03 (0.18)	0.03 (0.17)	0.03 (0.17)	0.03 (0.16)

Descriptive statistics reveal no obvious anomalies. As anticipated, mean years of education differ between the Census and CPS samples, largely because the Census measure is more coarsely defined, whereas the CPS captures recent attainment with greater precision. To determine whether these differences are statistically meaningful, we conduct formal difference-in-means tests across treatment and control groups, finding no significant differences.

Employment Protection & Education Decisions

Difference in means test: control vs. treatment groups (Census data)						
	Treated		Control		T-test	
	mean/b	sd/t	mean/b	sd/t	mean/b	sd/t
Education	8.61	1.34	8.46	1.29	-0.15***	(-36.59)
High_School	0.48	0.26	0.52	0.24	0.04***	(51.20)
College	0.41	0.25	0.38	0.23	-0.03***	(-45.94)
Age 15-19	0.18	0.38	0.18	0.39	0.00	(1.58)
Age 20-24	0.19	0.39	0.19	0.40	0.00	(0.32)
Age 25-29	0.20	0.40	0.20	0.40	-0.00	(-0.86)
Female	0.50	0.50	0.50	0.50	-0.00	(-1.87)
Black	0.26	0.44	0.30	0.46	0.04***	(30.81)
Asian	0.12	0.33	0.12	0.32	-0.00***	(-3.60)
Other race	0.30	0.46	0.28	0.45	-0.02***	(-11.09)
Observations	144019		333308		477327	

Difference in means test: control vs. treatment groups (CPS data)						
	Treated		Control		T-test	
	mean/b	sd/t	mean/b	sd/t	mean/b	sd/t
Education	7.67	2.30	7.39	2.36	-0.28***	(-64.91)
Wage99	18012.22	24199.63	16515.58	20603.20	-1496.64***	(-34.18)
Employed	0.80	0.40	0.79	0.41	-0.01***	(-17.59)
High School	0.56	0.50	0.63	0.48	0.08***	(82.57)
College	0.44	0.50	0.37	0.48	-0.08***	(-82.57)
Age	28.15	7.32	27.35	7.12	-0.81***	(-59.55)
Female	0.52	0.50	0.52	0.50	-0.00	(-0.56)
Black	0.10	0.30	0.14	0.35	0.04***	(73.87)
Asian	0.03	0.16	0.01	0.07	-0.02***	(-78.77)
Other ethnic group	0.03	0.17	0.02	0.15	-0.01***	(-18.83)
Observations	1479522		359132		1838654	

4.5 Empirical Framework

Our empirical strategy exploits a staggered Difference-in-Differences (DiD) design to address two questions: (1) how changes in dismissal costs affect returns to education (via labor-market outcomes), and (2) how those costs influence individuals' educational attainment decisions. Because each state adopted one or more of the three wrongful-discharge exceptions at different dates, we can use untreated states at a given time as controls. Throughout, Florida and Georgia serve as our common control group, having never enacted any exception during our sample period.

For the labor-market analysis, we focus on working-age respondents (ages 16–70), with a primary emphasis on the 16–40 cohort to ensure comparable experience profiles. Treatment cohorts are defined by age at interview (16–19, 20–24 and 25–29), while individuals aged 30–70 form the omitted (control) group. This grouping rests on the expectation that younger workers, lacking substantial experience, rely more on education as a signal following increases in firing costs.

For the labor-market analysis, we sample individuals aged 16–70 at the time of interview who reside in states that have enacted a firing-costs exception. To ensure comparable experience profiles, we concentrate on respondents aged 16–40, partitioning them into three treated cohorts (16–19, 20–24 and 25–29), and using those aged 30 and above as our control group. Younger workers are expected to rely more heavily on education as a signal once dismissal costs rise, since they have less experience for employers to observe.

All wage and earnings measures are deflated to 1999 real dollars. We capture educational attainment in two complementary ways: the first one is comprised of two binary indicators denoting (a) “high-school graduate or less” and (b) “some college or more;” the second one is formed by a continuous “years of schooling” measure, constructed following the CPS coding guidelines.

The specification we use, for individual i residing in state s at year t , is the following one:

$$Y_{ist} = \beta_0 Education_i + \beta_1 FC_{st} + \beta_2 Age_i + \beta_3 (Education_i \times FC_{st}) + \beta_4 (Education_i \times Age_i) + \beta_5 (Age_i \times FC_{st}) + \beta_6 (Education_i \times FC_{st} \times Age_i) + X_i + \alpha_s + \delta_t + \gamma_{c(i)} + \varepsilon_{ist} \quad (4.1)$$

Where:

- The dependent variable Y_{ist} can be either the individual’s reported real wage, its logarithm, or an employment-status dummy (1 if the respondent worked at least one week in the past year; 0 otherwise). We also repeat the analysis using total income (and its log), which yields similar but slightly less precise estimates.
- $Education_i$ is measured in two alternative ways:
 1. A continuous “years of schooling” variable ranging 0–12 (where 0 means high-school dropout and 12 means more than university degree education).
 2. A categorical variable taking values 0, 1 and 2, where 0 indicates high school dropout, 1 indicates at least high school but no college degree, and 2 indicates college degree or more.
- FC_{st} is specified either as separate dummies for each of the three firing-cost laws (C, G, P), as a single count variable (0–3) indicating how many laws are active in state s at time t or as a dummy variable indicating the presence of some firing-cost law.
- Age_i is defined as dummies for the age groups 15–19, 20–24 and 25–29. We also run specifications using age as a continuous variable.
- X_i is the vector of individual controls (gender, race, etc.).
- α_s are state-of-residence fixed effects.
- δ_t are year-of-interview fixed effects.
- $\gamma_{c(i)}$ are age-at-interview fixed effects.

We also include interactions among all fixed effects (state, year, and age) to capture any region- or cohort-specific temporal dynamics.

For the education-attainment analysis, we follow a similar identification strategy. In addition to using Florida and Georgia as treated units, we augment the control group with individuals from those same states who, at the time of the reform, were aged thirty or older and thus unlikely to alter their educational trajectories in response to the policy. We choose age thirty as our cutoff because, beyond that point, fewer than 10 percent of survey respondents report that they are still studying (see [Figure 4.2](#)). Finally, we expand our control set by including states whose discharge doctrine change occurred at least sixteen years after the treated states. This ensures that the youngest cohort in our sample (those aged fifteen at the time of the initial reform) will have reached age thirty by the time the control-state reform is implemented, thereby guaranteeing that all individuals in these later-reforming states remain unaffected by policy during our estimation window.

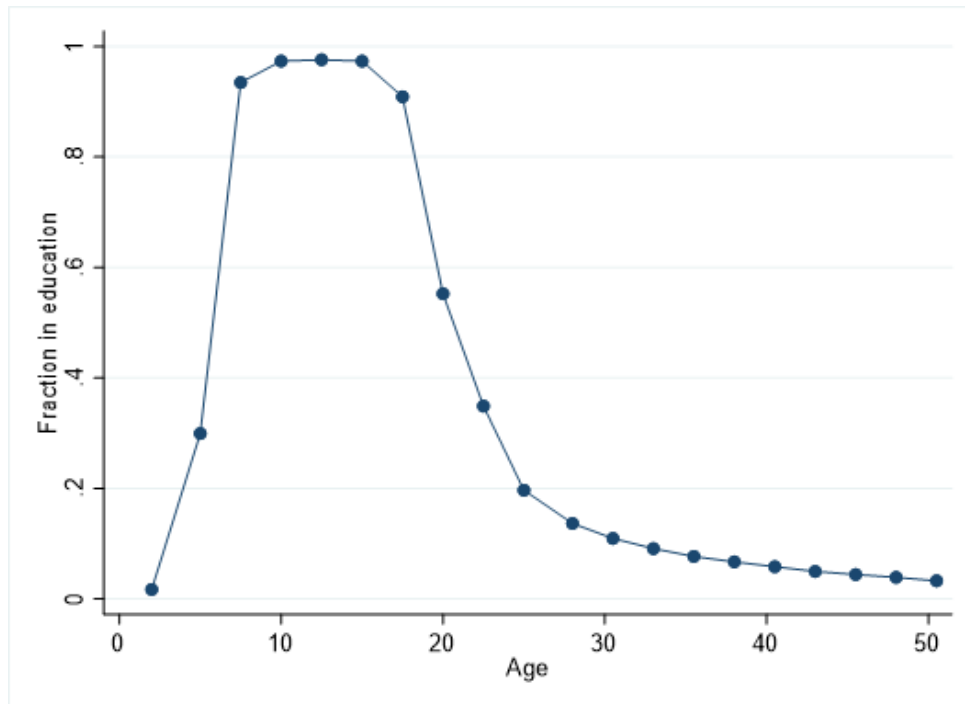


Fig. 4.2. Fraction of people reporting to be studying as their primary activity as a function of age.
Source: own elaboration using the data provided by the American Census

As said before, in [Figure 4.2](#) we observe that enrollment in schooling remains near 100 percent through age 18 and then declines sharply, falling below 10 percent by age 30. Accordingly, we partition our sample into one control group and three “treatment” cohorts, defined by age at the time the new doctrine was adopted in each individual’s birth state:

- **Control:** Respondents who were between 30 and 40 when their birth state adopted the reform: individuals are unlikely to alter their education level in response to the policy.
- **Group 1:** Those who were 15–19 at adoption: the cohort facing the mandatory-education cutoff exactly when the doctrine changed, and thus most able to adjust their ultimate attainment.
- **Group 2:** Those who were 20–24 at adoption: still young enough to extend their schooling but past the compulsory-education threshold, and therefore likely to respond, albeit to a lesser degree than Group 1.
- **Group 3:** Those who were 25–29 at adoption: many will already have completed their education, so the reform’s incentive to return to school is weakest here.

Taking all relevant factors into account, we define a group indicator variable, g , for each treated state. Each group g includes all observations from the treated state as well as from its corresponding control states. For example, group $g = 1$ comprises observations from Alabama (the treated state) and its control states, Florida and Georgia. Similarly, group $g = 10$ includes observations from Illinois (the treated state) and its control states: Delaware, Florida, and Georgia. This grouping structure allows for consistent comparison between treated and control units within each group, facilitating more accurate estimation of treatment effects. The baseline specification is the following:

$$\begin{aligned}
 Education_{csg} = & \alpha_{sg} + \delta_{cg} + \gamma_t + \beta_1(Group1_{cg} \times TreatedState_{sg}) + \beta_2(Group2_{cg} \times TreatedState_{sg}) \\
 & + \beta_3(Group3_{cg} \times TreatedState_{sg}) + X_{csg} + \epsilon_{csg}
 \end{aligned}
 \tag{4.2}$$

The dependent variable, $Education_{csg}$, captures the highest level of education reported by individuals defined by the aforementioned variables. The treatment indicator, $TreatedState_{sg}$, equals one for individuals born in the treated state within group g . Subscripts are defined as follows: c denotes the birth cohort, s the state of birth, and g the group of treated and control states. We include a set of control variables, including dummies for sex and race, as well as fixed effects for state-by-group, cohort-by-group, and year of interview, to account for potential regional and temporal confounders. The parameters of interest are β_1 , β_2 , and β_3 , which capture the estimated effects of the treatment across different specifications.

We conduct the analysis under multiple specifications. First, we define treatment as the adoption of any doctrine, regardless of type. Second, we distinguish between the adoption of each of the three specific doctrines. To assess robustness, we replicate the analysis using both the U.S. Census and the CPS datasets. Additionally, we estimate a cohort-by-cohort specification, defining a separate group for each age and using individuals aged 30 as the baseline. This specification takes the following form:

$$\begin{aligned}
 Education_{csg} = & \alpha_{sg} + \delta_{cg} + \gamma_t + \sum_{k=15}^{k=29} \beta_k(GroupK_{cg} \times TreatedState_{sg}) \\
 & + \sum_{k=31}^{k=40} \beta_k(GroupK_{cg} \times TreatedState_{sg}) + X_{csg} + \epsilon_{csg}
 \end{aligned}
 \tag{4.3}$$

4.6 Results and Discussion

4.6.1 Effects on the Labor Market

We start by assessing the impact of dismissal-cost reforms on the education wage premium. In order to do that, we first estimate a baseline earnings equation for our full sample. This model relates individual wages to the level of education, controlling for key covariates and allowing the coefficient on education to vary with the presence of employment-at-will exceptions. Formally, we specify:

$$Wage_{ist} = \beta_1 Education_i \times FC_{st} + X_{ist} + \epsilon_{ist}
 \tag{4.4}$$

Subsequently, we modify our baseline specification by introducing age-group indicators and their interactions with both education and the policy variables. Specifically, each regression includes dummies for our defined age cohorts (i.e. 15–19, 20–24, 25–29), the policy indicator, and the interactions between education, policy, and age groups. Columns in [Table 4.1](#) report four nested specifications.

In the first column we include only education level, the policy dummy, and our standard controls (gender and race), and year- and state-fixed effects. Here the policy indicator itself is small (−\$374, not significant), while each additional level of schooling raises annual wages by about \$12,590, indicating no average policy impact on wages once we account for education. In the second column we add the education by policy interaction: the significant negative effect on the policy dummy (−\$3,784) now captures the policy’s impact at the base level of education (high school dropout), while the positive interaction term (\$2,507.8 per education level) implies that the policy penalizes workers with very little schooling but rewards those with more education.

In column 3 we introduce age-group indicators (15–19, 20–24, and 25–29) along with their interactions with education. The age dummies show that base wages rise steeply through the life cycle, while the education by age interactions reveal that the return to an increase in the level of schooling varies by age group. Crucially, the education by policy slope (\$1,904.1) remains positive and significant, so even after netting out life-cycle patterns, the policy still accentuates the wage premium to education on average.

Finally, in column 4 we layer on the policy by age interactions, allowing the policy’s level effect to differ by cohort, and include the three-way education by policy by age terms. The return to an increase in the level of schooling under the policy jumps to about \$9,422 at ages 15–19, \$8,555 at 20–24, \$8,227 at 25–29, versus just \$2,638 for 30+, showing how high dismissal costs drive firms to lean most on formal education among younger cohorts.

Overall, these results paint a clear picture: higher dismissal costs lead firms to lean more heavily on formal education as a reliable signal of worker quality. Young workers with more education perceive substantially higher wages under these policies because their diplomas stand in for employers’ confidence that they will perform well. For older cohorts the education premium remains positive but is slightly smaller: firms still value schooling, but they also draw on workers’ accumulated experience records over multiple years. In this way, stringent firing regulations sharpen the sorting power of education, leading firms to pay up front for the credential that best distinguishes “good” from “bad” hires when letting someone go is costly.

Table 4.1: Effect of the presence of dismissal-cost policies on wages

	(1) Wage	(2) Wage	(3) Wage	(4) Wage
Education level	12 590.3*** (286.5)	10 551.6*** (703.5)	11 358.2*** (511.6)	10 660.0*** (807.4)
Policy dummy	−374.2 (300.6)	−3 784.0*** (999.6)	−3 606.6*** (771.0)	−14 180.2*** (1 455.7)
Education × policy		2 507.8*** (755.5)	1 904.2*** (509.0)	2 637.8*** (811.5)
Age 15–19			2 110.7** (1 016.1)	−8 466.3*** (1 426.5)
Age 20–24			5 135.4*** (925.8)	−2 517.6** (1 177.5)
Age 25–29			9 872.6*** (798.7)	4 145.5*** (1 273.1)
Educ × age15–19			309.1 (543.1)	−7 643.6*** (993.9)
Educ × age20–24			−1 422.6*** (399.1)	−8 799.1*** (846.0)
Educ × age25–29			−3 238.2*** (399.1)	−7 918.5*** (850.7)
Age15–19 × Policy				14 282.5*** (1 602.8)
Age20–24 × policy				13 725.2*** (1 326.3)
Age25–29 × policy				8 417.9*** (1 423.8)
Educ × age15–19 × policy				6 784.7*** (1 052.7)
Educ × age20–24 × policy				5 916.9*** (920.2)
Educ × age25–29 × policy				5 589.5*** (887.0)
Controls	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓
State fixed effects	✓	✓	✓	✓
Observations	1 838 654	1 838 654	1 838 654	1 838 654
Mean of Dependent Variable	17 719.9	17 719.9	17 719.9	17 719.9

Standard errors in parentheses, clustered by state of residence. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Wages in real \$ of 1999.

In [Table 4.2](#) we show the estimates for the impact of dismissal-cost doctrines on the probability of employment by regressing a binary employed indicator on level of schooling, policy dummies, their interactions, and controls. Because our data are cross-sectional (each individual appears only once), the dependent variable lacks within-person variation over time, which limits our ability to detect subtle shifts in employment status. Column (1) presents our benchmark: with only level of schooling, the policy

dummy, standard demographic controls, and year- and state-fixed effects, we see that each additional level of education raises the likelihood of being employed by roughly 14.4 percentage points, while the average effect of facing high dismissal costs is essentially zero and statistically indistinguishable from zero. In column (2) we introduce the interaction between education and the dismissal-cost dummy. Here the policy indicator at the baseline level of schooling is -3.7 pp, implying that for dropouts the burden of high firing costs significantly reduces hiring. However, increasing the level of schooling offsets that penalty by about 2.7 pp. Thus, even though firms with strict firing rules are more reluctant to hire the least-educated workers, they reward additional credentials more strongly.

In column (3) we introduce dummies for ages 15–19, 20–24 and 25–29 along with their interactions with education and the dismissal-cost dummy, which lets us see how both baseline employment and the return to each additional level of schooling vary across the life cycle when firing is costly; and in the final model in column (4) we allow the level effect of the policy and its schooling-return modulation to differ fully across our three age brackets. The negative coefficients on these three-way terms imply that the incremental benefit of education under dismissal costs diminishes somewhat for younger age groups compared to older ones. However, the overall pattern remains: younger workers with more education benefit from higher employment probabilities under dismissal costs, suggesting firms lean on formal education as a proxy for quality when firing is costly. We find that the net employment effect of high dismissal costs is most positive for the very youngest (15–19), slightly positive for ages 20–24, and approximately zero for ages 25–29, relative to a -17.0 pp drop for the 30+ reference group. Meanwhile, the three-way interactions reveal that under strict firing rules the youngest cohort actually sees a reduction in the marginal benefit of an extra level of schooling (-0.179 pp), whereas the negative three-way terms for the 20–24 and 25–29 groups (-0.133 pp and -0.080 pp, respectively) indicate a smaller education penalty under policy for older youth. In particular, the more negative three-way interaction for the 15–19 group likely reflects that additional years of schooling in this cohort often substitute for, rather than complement, on-the-job experience; when both experience and education serve as signals, high dismissal costs may lead firms to discount very young workers who lack a minimum threshold of work tenure despite their formal credentials.

Taken together, these nested specifications show that higher dismissal costs make firms more cautious about hiring low-education or inexperienced applicants, driving down their employment probabilities, while simultaneously amplifying the value of formal schooling as a hiring signal, especially for those with moderate levels of education and some work history.

To further assess the robustness of our findings, we re-estimate the models in [Table 4.1](#) by replacing the categorical education level variable with the continuous measure of years of schooling (ranging from 0 to 12). The results, presented in [Table A4.1](#) of the Appendix, mirror our main findings, confirming that higher dismissal costs continue to exert an effect on the wage returns to additional years of schooling.

Table 4.2: Effect of the presence of dismissal-cost policies on employment

	(1) Employed	(2) Employed	(3) Employed	(4) Employed
Education level	0.144*** (0.00539)	0.123*** (0.00779)	0.124*** (0.00643)	0.105*** (0.0134)
Dismissal Cost Dummy	-0.000165 (0.00483)	-0.0366*** (0.0106)	-0.136*** (0.0167)	-0.170*** (0.0262)
Education × Dismissal Cost		0.0268*** (0.00645)	0.0500*** (0.00613)	0.0745*** (0.0139)
Age 15–20			-0.154*** (0.0232)	-0.335*** (0.0801)
Age 20–25			0.0290* (0.0152)	-0.111*** (0.0297)
Age 25–30			0.107*** (0.0121)	0.0237 (0.0242)
Education × Age 15–20			-0.0659*** (0.00304)	0.107 (0.0691)
Education × Age 20–25			-0.0727*** (0.00342)	0.0422** (0.0157)
Education × Age 25–30			-0.0726*** (0.00435)	-0.0124 (0.0141)
Age 15–20 × Dismissal Cost			0.300*** (0.0239)	0.489*** (0.0814)
Age 20–25 × Dismissal Cost			0.137*** (0.0182)	0.303*** (0.0357)
Age 25–30 × Dismissal Cost			0.0570*** (0.0124)	0.168*** (0.0289)
Education × Age 15–20 × Dismissal Cost				-0.179** (0.0691)
Education × Age 20–25 × Dismissal Cost				-0.133*** (0.0167)
Education × Age 25–30 × Dismissal Cost				-0.0796*** (0.0147)
Controls	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓
State fixed effects	✓	✓	✓	✓
Observations	1,838,654	1,838,654	1,838,654	1,838,654
Mean of Dependent Variable	0.797	0.797	0.797	0.797
Standard errors in parentheses, clustered by state of residence.			* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$	

Robustness checks

We next carry out a robustness check by analyzing the three wrongful-discharge laws separately, which will allow us to identify which doctrine exerts the strongest influence on education returns. To investigate this, we estimate our baseline wage and employment regressions separately for each exception (implied contract, public policy, and good faith) using distinct state-level indicators for their adoption. A small summary of the results is provided in Table ?? below, and the full set of results appears in Table A4.2 and Table A4.3 in the appendix.

Table 4.3: Effect of each specific dismissal-cost doctrine on wages and employment

	(1) Wage	(2) Employed
Education level	11 140.3*** (576.1)	0.121*** (0.0109)
Policy G	−2 996.0* (1 646.6)	−0.00602 (0.0285)
Policy C	−5 821.5*** (1 872.8)	−0.0778** (0.0343)
Policy P	−7 150.7*** (2 097.4)	−0.0584 (0.0410)
Educ. × C	1 187.9** (530.2)	0.0394** (0.0178)
Educ. × P	1 119.1* (613.2)	0.0240 (0.0223)
Age 15–19 × C	9 837.0*** (3 193.7)	0.236*** (0.0641)
Age 15–19 × P	13 525.4*** (2 354.5)	0.261*** (0.0527)
Educ. × Age 20–24 × C	1 959.2* (1 100.6)	−0.0684*** (0.0212)
Educ. × Age 25–29 × C	3 340.7*** (971.3)	−0.0396*** (0.0125)
Year fixed effects	✓	✓
State fixed effects	✓	✓
Observations	1 838 654	1 838 654
Mean dep. var.	17 719.9	0.797

Standard errors in parentheses, clustered by state. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

For the standard wage specification, the implied contract exception yields the largest and most robust effect in the schooling premium, while the public policy exception also produces a modest effect. By contrast, the good faith exception generates small, statistically insignificant coefficients in every model. This latter result may reflect its lack of adoption: only twelve states enacted the good faith doctrine during our sample period, whereas virtually all but seven jurisdictions adopted both the implied contract and public policy exceptions. When we repeat these doctrine-specific estimations using the employment-probability outcome, the pattern remains unchanged. Both the implied contract and public policy exceptions are associated with meaningful, though not always statistically significant, effects in the education premium, while the good faith exception continues to show no discernible effect.

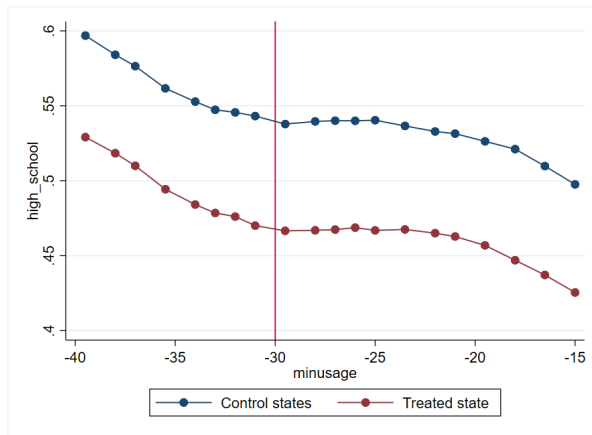
To assess the robustness of our findings, we conduct two additional exercises. First, in table [Table A4.4](#) in the Appendix, we broaden the control group to include all individuals aged thirty to seventy, rather than restricting it to those between thirty and forty. The key policy coefficients and their education and age interactions are somewhat smaller in magnitude and, in a few cases, marginally less precisely estimated, but their signs and statistical significance endure, suggesting the results are not driven by an idiosyncratic choice of control cohort. The main effect of education is likewise stable. Finally, in table [Table A4.5](#) in the Appendix, we use total personal income rather than wage earnings as the dependent variable. Despite increased noise from non-wage income sources, the education–policy interaction retains its magnitude and significance, closely mirroring the results obtained with wage income. Together, these robustness checks exhibit that our main conclusions remain unchanged.

4.6.2 Effects on educational attainment

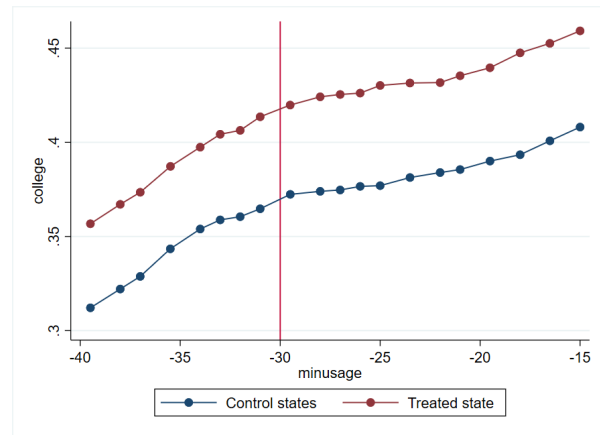
At least one new doctrine adopted

We now turn to the impact of wrongful-discharge laws on individuals' educational attainment (that is, how much schooling they choose to complete). To do so, we employ a difference-in-differences approach in which the treatment is defined simply as the adoption of at least one of the new dismissal-cost doctrines, irrespective of which specific exception is in force or how many have been adopted. A key identifying assumption in this framework is that treated and control groups would have followed parallel trends in education in the absence of the policy change. To assess this, we construct [Figure 4.3](#), showing education level by age using two binary outcomes: not attending college (i.e. finishing high school or being a dropout) and going to college (completing a degree or at least some years of college). The Y axis will display the percentage of people with the respective level of education, while the X axis will exhibit the age of the respondent (in negative values to facilitate reading). The left-hand panel shows the proportion of individuals with a maximum education level of high-school diploma or less, while the right-hand panel displays the same series for those with at least some year of college. In both cases, the pre-intervention trends for high school completion and college attendance for both control and treatment groups track one another closely, confirming the parallel-trends assumption.

Although we observe that treated states begin with a marginally higher share of college-educated individuals (and, correspondingly, a slightly lower share with only a high school diploma), these level differences are small in magnitude and remain stable over the pre-treatment period. Because these gaps are constant, they are absorbed by the model's fixed effects and do not threaten the validity of our difference-in-differences estimates.



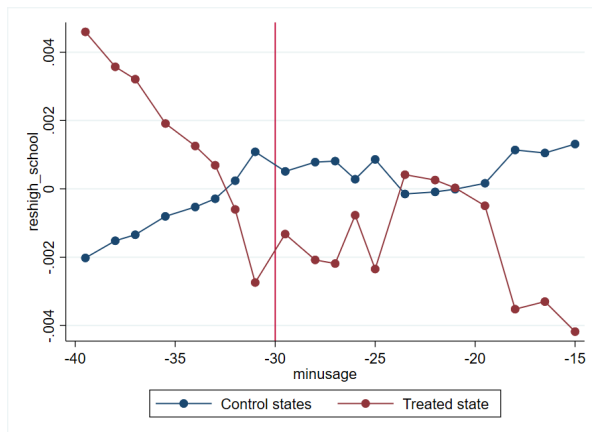
Percentage High School education or less.



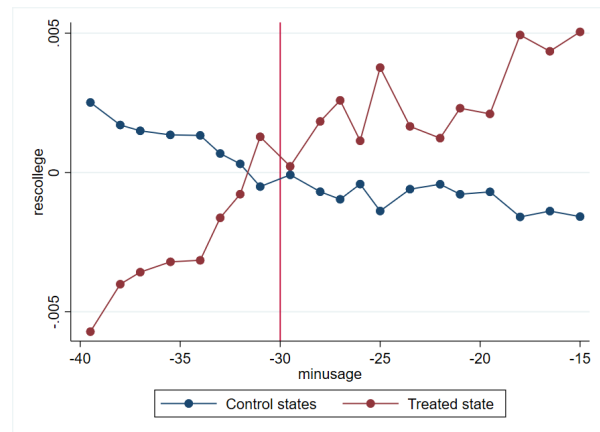
Percentage College education or more.

Fig. 4.3. Education trends across control and treated states.
Source: own elaboration using the data provided by the American Census

We extend our analysis by examining whether doctrine adoption alters overall educational attainment after accounting for observable characteristics. To do so, we estimate our baseline educational regression, controlling for demographics, cohort, and state fixed effects; and then plot the residuals (i.e. actual minus predicted education) against treatment status. As shown in [Figure 4.4](#), the average residual in treated states lies marginally above that in control states, indicating a small positive adjustment in schooling following policy adoption. However, this difference fails to reach statistical significance, suggesting that, once covariates are held constant, the introduction of dismissal-cost doctrines did not meaningfully change individuals' completed years of education. Results for this regression are shown in [Table 4.4](#).



Percentage of residuals, High School education or less.



Percentage of residuals, College education or more.

Fig. 4.4. Average residual of education attainment across control and treated states.
Source: own elaboration using the data provided by the American Census

Table 4.4

	(1) Years of Education	(2) Education level	(3) High School dummy	(4) College dummy
Treated State $\times$ Age 15-19	-0.0917 (0.124)	-0.0328 (0.125)	-0.00707 (0.0221)	0.0108 (0.0168)
Treated State $\times$ Age 20-24	-0.105 (0.0989)	-0.0492 (0.105)	-0.00205 (0.0176)	0.00704 (0.0124)
Treated State $\times$ Age 25-29	-0.0676 (0.0670)	-0.0255 (0.0713)	-0.00490 (0.0111)	0.00711 (0.00709)
Year-of-survey FE	✓	✓	✓	✓
State-by-group FE	✓	✓	✓	✓
Cohort-by-group FE	✓	✓	✓	✓
Observations	477327	477327	477327	477327
Mean of Dep. Variable	8.628	7.092	0.526	0.380

Standard errors in parentheses, clustered by state of birth. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Overall, the adoption of at-will exceptions appears to have a negligible impact on individuals' educational attainment. When we estimate the effect on completed years of schooling (whether measured continuously or as a categorical high- vs. low-education variable) the coefficients are consistently negative but both statistically insignificant and very small in magnitude. Conversely, when we analyze the likelihood of finishing high school versus enrolling in college, the point estimates flip sign (indicating a small increase in college attendance and a slight decrease in high-school-only attainment), yet these effects remain vanishingly small and never reach significance. Taken together, these findings suggest that the introduction of dismissal-cost doctrines did not meaningfully alter how long people stayed in school.

To ensure these conclusions are robust, we implement a series of additional checks, like analyzing each of the three discharge-laws separately and using a cohort-by-cohort specification.

Robustness checks

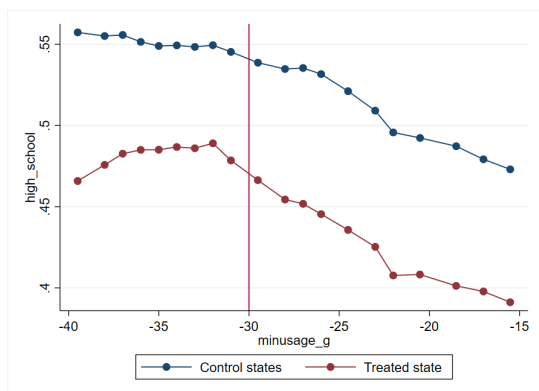
We extend our analysis by estimating our difference-in-differences model separately for each of the three dismissal-cost doctrines: implied contract, public policy, and good faith; using a distinct policy indicator for each. Graphical inspection (shown in in [Figure 4.5](#)) shows the same trends for the implied contract and public policy exceptions, while the good faith exception shows a slight divergence likely driven by its late and uneven adoption across only a handful of states. In all three cases, treated states begin with education levels roughly five percentage points higher than controls, a gap that remains stable over time.

In [Table 4.5](#), columns 1–3 present estimates from separate regressions in which the Good Faith, Public Policy, and Implied Contract doctrines each serve as the treatment indicator. In column 4, we replace the binary treatment indicators with a variable indicating the presence of some doctrine, independently of which one or how many of them. In all specifications, the coefficients remain close to zero and fail to achieve statistical significance, indicating that no single exception by itself or the presence of some exception influenced individuals' educational choices.

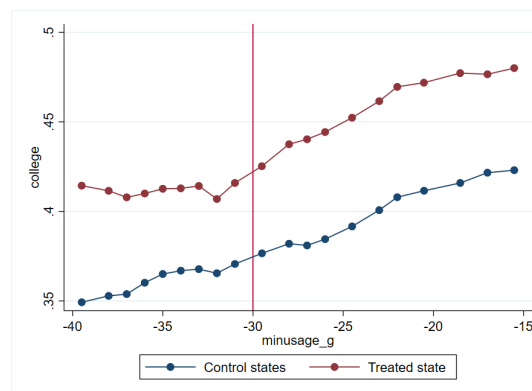
Table 4.5: Effect of the presence of each doctrine on educational achievement

	(1) Good Faith	(2) Public Policy	(3) Implied Contract	(4) Any Doctrine
Treated State $\times$ Age 15–19	–0.0745 (0.104)	–0.0439 (0.120)	–0.0889 (0.130)	–0.0467 (0.103)
Treated State $\times$ Age 20–24	–0.0878 (0.0760)	–0.0586 (0.0980)	–0.0826 (0.109)	–0.0506 (0.0706)
Treated State $\times$ Age 25–29	–0.0256 (0.0449)	–0.0354 (0.0669)	–0.0613 (0.0775)	–0.0433 (0.0559)
Year FE	✓	✓	✓	✓
State-by-group FE	✓	✓	✓	✓
Cohort-by-group FE	✓	✓	✓	✓
Observations	181,812	397,442	450,289	560,733
Mean of Dep. Var.	7.178	7.106	7.086	7.103
Dependent variable is the (average) years of education				
Standard errors in parentheses, clustered by state of birth.			* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.	

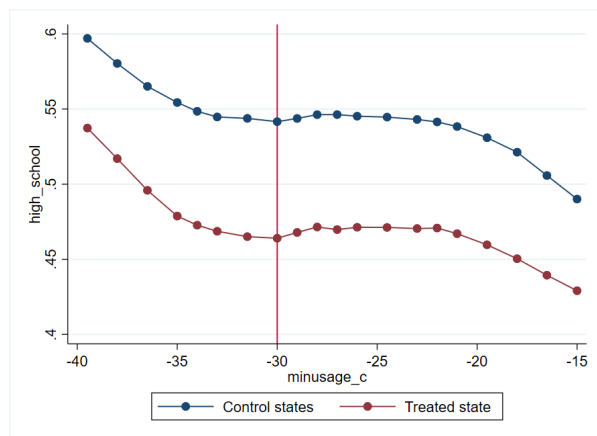
To conclude the robustness checks, we replicate the main estimation using the state of residence instead of the state of birth as indicator, and using an alternative dataset, the Current Population Survey (CPS), which provides comparable variables but draws from a different population sample and also uses state of residence rather than state of birth. The results, presented in Tables [Table A4.6](#) and [Table A4.7](#) in the appendix, remain consistent across both specifications: estimated effects are small, statistically insignificant, and retain the same sign as in the main analysis using the Census data.



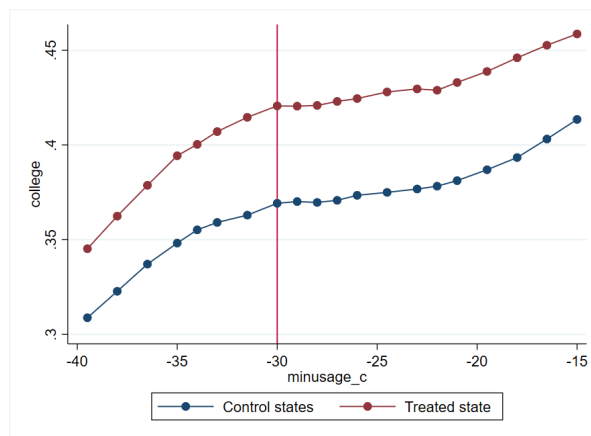
Percentage of High School education or less - Good faith doctrine



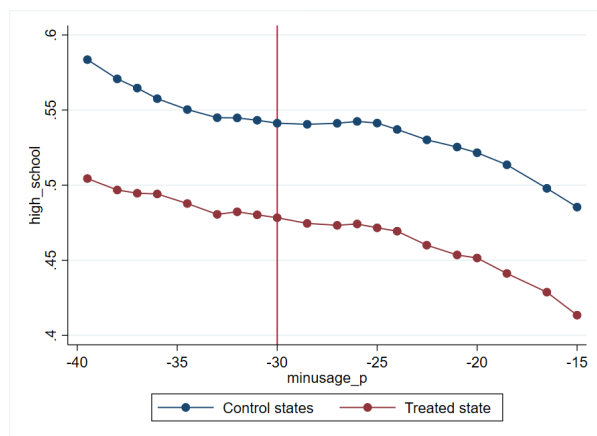
Percentage of College education or more - Good faith doctrine



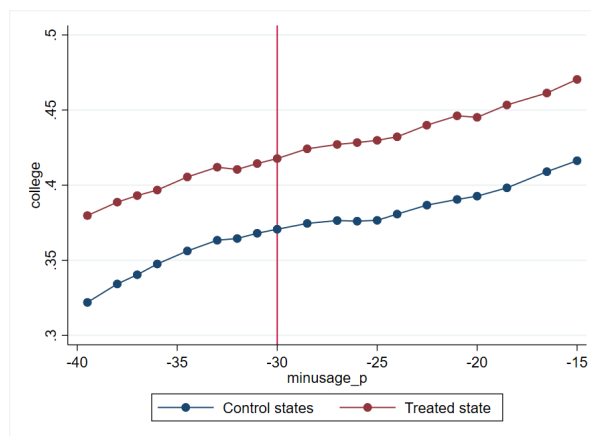
Percentage of High School education or less - Implied Contract doctrine



Percentage of College education or more - Implied Contract doctrine



Percentage of High School education or less - Public Policy doctrine



Percentage of College education or more - Public Policy doctrine

Fig. 4.5. Education trends across control and treated states.
Source: own elaboration using the data provided by the American Census

4.7 Summary and Concluding Remarks

This paper presents new evidence on how the gradual introduction of wrongful-discharge protections, through the implied contract, public policy, and good-faith exceptions to the at-will employment doctrine in the U.S., shaped both the returns to education in the labor market and individuals' educational attainment decisions. Drawing on a staggered differences-in-differences design, we exploit state-level variation in the timing of each doctrine's adoption between 1970 and 1999, using rich data from the U.S. Census and the Current Population Survey (CPS). By focusing on cohorts aged fifteen to forty at the time of each reform, we are able to isolate how the introduction of firing constraints affected the returns to schooling for workers at different stages of their career, as well as whether these evolving returns in turn reshaped individuals' own educational decisions.

Our core finding is that higher dismissal costs amplify the wage premium to formal schooling, especially for younger workers. In specifications in which level of education, presence of the policy and age groups interact between one another, we estimate that, relative to the over-thirty cohort, the extra wage return to a year of schooling under any of the three doctrines is two to three times larger for workers aged fifteen to twenty-four. This pattern holds for both the implied-contract and the public-policy exceptions and, to a smaller extent, for the good-faith exception in jurisdictions that adopted it. These results are robust to an array of fixed effects and alternative cohort-of-control definitions. The increase in the schooling premium under higher dismissal costs is consistent with a signaling interpretation: when firing becomes more difficult and expensive, employers compensate by placing greater reliance on observable credentials to screen and sort potential employees.

Despite these pronounced changes in market rewards, we find virtually no evidence that individuals adjusted their own schooling decisions in response. Difference-in-differences estimates of the education levels of individuals yield coefficients centered near zero and statistically indistinguishable from no effect, whether we define treatment as exposure to any doctrine or disaggregate by exception. Even among cohorts most affected by the rising returns on wages (the youngest workers), there is no detectable increase in educational attainment. This suggests that, although labor-market institutions can reshape the value of education, such signaling-driven returns alone do not appear to influence additional human-capital investment.

Turning to employment outcomes, the overall presence of dismissal-cost protections has only a modest net effect on participation, but it does alter the distribution of opportunities. Workers with low levels of schooling faced slight declines in employment probability after reform, whereas those with more education experienced small gains. The result is again strongest among younger cohorts, for whom the net effect is slightly positive, while older workers see little to no change. In tandem with the wage results, this pattern reinforces the notion that increase in firing restrictions induce firms to tilt hiring toward candidates with stronger observable credentials, raising wages for the educated but making it incrementally harder for the least-credentialed to enter in the workforce.

Taken together, our evidence paints a coherent picture in which rising dismissal costs lead employers to substitute toward more educated workers. This substitution drives up the wage returns to schooling without prompting workers to extend their education. The absence of a schooling response may reflect liquidity constraints, delayed information flows, or the fact that many of the affected cohorts were already past the prime years for formal education when reforms occurred.

From a policy standpoint, our findings highlight a potential trade-off inherent in labor-market regulation. Enhanced job security through stricter wrongful-discharge protections benefits incumbent workers, particularly the educated, by raising their wages. Yet, it may also exacerbate barriers for those with fewer credentials without compelling them to obtain more schooling. Policymakers seeking to strengthen

worker protections should therefore consider complementary measures, such as expanded access to adult education, targeted training vouchers, or tuition assistance, to ensure that disadvantaged groups can take advantage of the heightened returns to education. Moreover, because our results indicate that protective doctrines differ little in their labor-market impact once any form of dismissal cost is in place, legislative efforts might focus on the broad design of termination rules rather than the nuances of each exception.

Our study also points toward several avenues for future research. First, examining longer-run career trajectories could reveal whether the immediate wage and employment shifts we document translate into persistent lifetime earnings effects. Second, exploring heterogeneity across industries or occupations may uncover contexts in which schooling responses are more or less likely, or returns may change in sign or magnitude. Third, integrating measures of non-wage job quality, such as training opportunities or hours stability, would deepen our understanding of how employment protections shape overall worker welfare.

These findings underscore the power of labor-market institutions to alter the signaling value of credentials, with important implications for inequality and policy design in the modern economy.

4.8 Appendix

Table A4.1: Effect of the presence of dismissal–cost policies on wages, using years of schooling

	(1) Wage	(2) Wage	(3) Wage	(4) Wage
Education	3 713.0*** (72.08)	2 926.7*** (187.4)	2 760.9*** (246.5)	2 836.9*** (264.8)
Policy dummy	–346.5 (276.9)	–7 674.1*** (1 384.6)	–18 812.0*** (2 075.3)	–18 048.4*** (2 301.9)
Educ. × Policy		988.8*** (190.7)	1 355.4*** (231.1)	1 254.6*** (263.1)
Age 15–19			–10 279.3*** (1 165.8)	–3 582.5* (2 121.0)
Age 20–24			–4 176.3*** (939.1)	–2 113.4 (2 090.7)
Age 25–29			2 963.0*** (963.9)	5 663.2** (2 198.4)
Educ. × Age 15–19			–426.2*** (102.3)	–1 741.0*** (265.0)
Educ. × Age 20–24			–1 129.4*** (86.72)	–1 437.5*** (267.0)
Educ. × Age 25–29			–1 244.8*** (108.1)	–1 601.3*** (274.1)
Age 15–19 × Policy			16 973.0*** (932.0)	9 985.6*** (2 295.8)
Age 20–24 × Policy			18 346.4*** (584.3)	15 858.1*** (2 199.4)
Age 25–29 × Policy			14 131.9*** (392.9)	10 581.5*** (2 248.9)
Educ. × Age 15–19 × Policy				1 353.2*** (287.6)
Educ. × Age 20–24 × Policy				363.1 (276.4)
Educ. × Age 25–29 × Policy				463.4* (269.8)
Year fixed effects	✓	✓	✓	✓
State fixed effects	✓	✓	✓	✓
Observations	1 838 654	1 838 654	1 838 654	1 838 654
Mean of Dep. Variable	17 719.9	17 719.9	17 719.9	17 719.9

Standard errors in parentheses, clustered by state of residence. Wages in real \$ of 1999.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4.2: Effect of each of the possible policies on wages

	(1) wage	(2) wage	(3) wage	(4) wage
Education level	12 587.0*** (286.8)	10 652.8*** (580.9)	10 490.4*** (693.4)	11 140.3*** (576.1)
Policy G	−888.0* (509.2)	−1 852.4 (1 180.9)	−2 837.4* (1 587.1)	−2 996.0* (1 646.6)
Policy C	−57.0 (272.2)	−2 490.7*** (918.5)	−7 102.3*** (2 138.9)	−5 821.5*** (1 872.8)
Policy P	−400.8 (242.5)	−1 853.3* (1 037.0)	−7 545.3*** (2 617.6)	−7 150.7*** (2 097.4)
Educ. × G		673.7 (919.5)	746.9 (946.3)	1 018.0 (982.3)
Educ. × C		1 726.5** (692.8)	2 004.4** (758.9)	1 187.9** (530.2)
Educ. × P		1 027.8 (795.5)	1 321.0 (969.7)	1 119.1* (613.2)
Age 15–19			−9 724.9*** (1 995.7)	−14 078.8*** (2 264.4)
Age 20–24			−7 388.5*** (732.9)	−2 671.1** (1 182.9)
Age 25–29			639.9 (937.0)	4 979.6*** (1 083.9)
Educ. × Age 15–19			−1 201.6* (638.3)	2 177.0 (2 550.3)
Educ. × Age 20–24			−3 554.6*** (348.7)	−5 294.6*** (1 335.4)
Educ. × Age 25–29			−3 539.6*** (419.0)	−6 583.6*** (717.7)
Age 15–19 × G			1 639.0 (1 230.6)	4 402.9** (1 857.4)
Age 15–19 × C			8 687.5*** (2 340.1)	9 837.0*** (3 193.7)

Employment Protection & Education Decisions

Age 15–19 × P	9 890.7*** (2 928.9)	13 525.4*** (2 354.5)
Age 20–24 × G	1 294.8 (1 003.2)	3 299.8** (1 588.0)
Age 20–24 × C	6 183.2*** (1 634.2)	6 588.9*** (1 999.9)
Age 20–24 × P	7 257.9*** (2 261.0)	8 711.1*** (1 755.8)
Age 25–29 × G	2 863.4** (1 161.9)	1 082.2 (1 335.6)
Age 25–29 × C	7 481.7*** (1 766.0)	2 650.0** (1 235.7)
Age 25–29 × P	8 227.5*** (1 949.0)	5 802.5*** (1 265.9)
Educ. × Age 15–19 × G		–1 920.2** (937.3)
Educ. × Age 15–19 × C		–929.0 (1 575.0)
Educ. × Age 15–19 × P		–2 733.4 (1 868.4)
Educ. × Age 20–24 × G		–636.3 (1 114.7)
Educ. × Age 20–24 × C		1 959.2* (1 100.6)
Educ. × Age 20–24 × P		1 279.3 (1 308.9)
Educ. × Age 25–29 × G		1 144.7 (1 218.8)
Educ. × Age 25–29 × C		3 340.7*** (971.3)
Educ. × Age 25–29 × P		1 672.0 (1 013.4)
Year fixed effects	✓	✓

State fixed effects	✓	✓	✓	✓
Observations	1 838 654	1 838 654	1 838 654	1 838 654
Mean of Dependent Variable	17 719.9	17 719.9	17 719.9	17 719.9

G, C and P refer to each of the possible dismissal-cost doctrines: the Good Faith exception, the Implied Contract exception and the Public Policy exception, respectively.

Standard errors in parentheses, clustered by state of residence * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4.3: Effect of each of the possible policies on employment probabilities

	(1) Employed	(2) Employed	(3) Employed	(4) Employed
Education level	0.144*** (0.00538)	0.133*** (0.00974)	0.149*** (0.0100)	0.121*** (0.0109)
Policy G	−0.00104 (0.00718)	0.00740 (0.0171)	−0.00392 (0.0228)	−0.00602 (0.0285)
Policy C	−0.000746 (0.00492)	−0.0277* (0.0156)	−0.0534** (0.0253)	−0.0778** (0.0343)
Policy P	−0.000260 (0.00448)	0.00171 (0.0233)	−0.0250 (0.0357)	−0.0584 (0.0410)
Educ. × G		−0.00576 (0.00866)	−0.00479 (0.00888)	−0.00314 (0.0129)
Educ. × C		0.0192* (0.0103)	0.0217* (0.0117)	0.0394** (0.0178)
Educ. × P		−0.00157 (0.0164)	0.000801 (0.0189)	0.0240 (0.0223)
Age 15–19			−0.0120 (0.0202)	−0.275*** (0.0386)
Age 20–24			0.0453*** (0.0131)	−0.0689*** (0.0198)
Age 25–29			0.124*** (0.0109)	0.0485*** (0.0180)
Educ. × Age 15–19			−0.0565*** (0.00615)	0.149*** (0.0187)
Educ. × Age 20–24			−0.0682*** (0.00322)	0.0386*** (0.00955)

Employment Protection & Education Decisions

Educ. $\times$ Age 25–29	–0.0725*** (0.00354)	–0.0188** (0.00840)
Age 15–19 $\times$ G	0.0266* (0.0146)	0.0599 (0.0403)
Age 15–19 $\times$ C	0.0789*** (0.0246)	0.236*** (0.0641)
Age 15–19 $\times$ P	0.0819*** (0.0249)	0.261*** (0.0527)
Age 20–24 $\times$ G	0.00944 (0.0123)	0.0446 (0.0379)
Age 20–24 $\times$ C	0.0140 (0.0113)	0.139*** (0.0459)
Age 20–24 $\times$ P	0.0118 (0.0144)	0.153*** (0.0410)
Age 25–29 $\times$ G	0.0183 (0.0137)	0.0316 (0.0326)
Age 25–29 $\times$ C	0.0221* (0.0120)	0.0783*** (0.0283)
Age 25–29 $\times$ P	0.0164 (0.0144)	0.0828*** (0.0302)
Educ. $\times$ Age 15–19 $\times$ G		–0.0248 (0.0199)
Educ. $\times$ Age 15–19 $\times$ C		–0.120*** (0.0327)
Educ. $\times$ Age 15–19 $\times$ P		–0.136*** (0.0260)
Educ. $\times$ Age 20–24 $\times$ G		–0.0175 (0.0176)
Educ. $\times$ Age 20–24 $\times$ C		–0.0684*** (0.0212)
Educ. $\times$ Age 20–24 $\times$ P		–0.0792*** (0.0180)

Educ. $\times$ Age 25–29 $\times$ G				–0.00894 (0.0145)
Educ. $\times$ Age 25–29 $\times$ C				–0.0396*** (0.0125)
Educ. $\times$ Age 25–29 $\times$ P				–0.0464*** (0.0123)
Year fixed effects	✓	✓	✓	✓
State fixed effects	✓	✓	✓	✓
Observations	1 838 654	1 838 654	1 838 654	1 838 654
Mean of Dependent Variable	0.797	0.797	0.797	0.797

G, C and P refer to each of the possible dismissal-cost doctrines: the Good Faith exception, the Implied Contract exception and the Public Policy exception, respectively.

Standard errors in parentheses, clustered by state of residence * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4.4: Using 30-70 years old as baseline treatment group

	(1) Wage	(2) Wage	(3) Wage	(4) Wage
Education level	14 486.6*** (335.7)	12 010.5*** (731.1)	11 522.0*** (885.3)	12 789.9*** (590.6)
Policy dummy	-424.6 (306.3)	-4 409.9*** (1 005.5)	-9 123.4*** (1 219.6)	-6 974.3*** (882.9)
Education × policy		3 010.0*** (796.7)	3 070.6*** (916.8)	1 446.3** (634.5)
Age 15–19			-14 993.0*** (705.5)	-4 358.3*** (1 281.3)
Age 20–24			-10 564.7*** (577.2)	1 369.7 (866.3)
Age 25–29			-4 281.1*** (492.1)	8 173.4*** (1 032.4)
Education × Age 15–19			640.6** (263.0)	-9 595.4*** (942.6)
Education × Age 20–24			-789.2*** (278.8)	-10 745.3*** (703.1)
Education × Age 25–29			-909.4*** (316.5)	-10 034.1*** (694.7)
Age 15–19 × policy			17 372.1*** (590.2)	6 221.2*** (1 238.6)
Age 20–24 × policy			17 666.3*** (462.3)	4 123.1*** (950.1)
Age 25–29 × policy			12 853.3*** (327.9)	-2 048.5* (1 115.6)
Education × Age 15–19 × policy				10 602.2*** (909.1)
Education × Age 20–24 × policy				11 044.5*** (811.6)
Education × Age 25–29 × policy				10 759.1*** (762.6)
Year fixed effects	✓	✓	✓	✓
State fixed effects	✓	✓	✓	✓
Observations	3 354 202	3 354 202	3 354 202	3 354 202
Mean of Dep. Variable	19 834.3	19 834.3	19 834.3	19 834.3

Standard errors in parentheses, clustered by state of residence. Wages in real \$ of 1999.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4.5: Using total personal income instead of wage earnings

	(1) Total Income	(2) Total Income	(3) Total Income	(4) Total Income
Education level	16 632.7*** (369.8)	13 969.7*** (858.7)	13 438.5*** (1 034.7)	14 819.6*** (715.5)
Policy dummy	−436.0 (353.9)	−4 722.1*** (1 224.8)	−10 623.1*** (1 487.0)	−8 280.7*** (1 113.8)
Education × Policy		3 237.1*** (912.8)	3 807.9*** (1 061.3)	2 038.4*** (757.5)
Age 15–19			−18 419.4*** (734.2)	−7 305.9*** (1 532.4)
Age 20–24			−13 846.8*** (638.9)	−1 612.8 (1 033.1)
Age 25–29			−7 523.7*** (513.1)	6 547.1*** (1 169.3)
Education × Age 15–19			−382.6 (327.3)	−11 062.5*** (1 158.6)
Education × Age 20–24			−1 730.5*** (368.0)	−11 928.0*** (841.4)
Education × Age 25–29			−1 521.1*** (383.2)	−11 827.8*** (835.2)
Age 15–19 × Policy			19 597.9*** (670.6)	7 922.7*** (1 542.7)
Age 20–24 × Policy			20 353.5*** (507.9)	6 449.5*** (1 145.3)
Age 25–29 × Policy			15 787.7*** (319.3)	−1 039.9 (1 314.0)
Education × Age 15–19 × Policy				11 078.3*** (1 149.9)
Education × Age 20–24 × Policy				11 328.6*** (965.6)
Education × Age 25–29 × Policy				12 146.6*** (927.0)
Year fixed effects	✓	✓	✓	✓
State fixed effects	✓	✓	✓	✓
Observations	3 354 202	3 354 202	3 354 202	3 354 202
Mean of dependent variable	24 709.0	24 709.0	24 709.0	24 709.0

Standard errors in parentheses, clustered by state of residence. Income in real \$ of 1999.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4.6: Using State of Residence in the Census

	1	2	3	4
	Years of Education	Education level	High School dummy	College dummy
Treated State $\times$ Age 15-19	0.104 (0.0843)	0.0286 (0.0914)	-0.0215 (0.0227)	0.0222 (0.0165)
Treated State $\times$ Age 20-24	0.0736 (0.0350)	0.0271 (0.0679)	-0.0122 (0.0161)	0.0102 (0.0139)
Treated State $\times$ Age 25-29	-0.0427 (0.0278)	0.0136 (0.0443)	0.000205 (0.0103)	0.000803 (0.00907)
Year-of-survey FE	✓	✓	✓	✓
State-by-group FE	✓	✓	✓	✓
Cohort-by-group FE	✓	✓	✓	✓
Observations	11950	11950	11950	11950
Mean of Dep. Variable	8.770	7.240	0.484	0.412
Standard errors in parentheses				
Std. errors clustered by state of residence				
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$				

Table A4.7: Using State of Residence in the CPS

	1	2	3	4
	Years of Education	Education level	High School dummy	College dummy
Treated State $\times$ Age 15-19	0.0292 (0.0728)	0.0293 (0.0182)	-0.0298 (0.0180)	0.0298 (0.0180)
Treated State $\times$ Age 20-24	0.0252 (0.0423)	0.00831 (0.00798)	-0.00771 (0.00793)	0.00771 (0.00793)
Treated State $\times$ Age 25-29	-0.00844 (0.0361)	0.000984 (0.00847)	-0.000893 (0.00831)	0.000893 (0.00831)
Year-of-survey FE	✓	✓	✓	✓
State-by-group FE	✓	✓	✓	✓
Cohort-by-group FE	✓	✓	✓	✓
Observations	3234447	3234447	3234447	3234447
Mean of Dep. Variable	8.280	1.546	0.451	0.549
Standard errors in parentheses				
Std. errors clustered by state of residence				
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$				

Table A4.8: Year of adoption of each of the three possible dismissal costs doctrines by each U.S. state

Policy	C			P			G		
State	Pre	Year of adoption	Post	Pre	Year of adoption	Post	Pre	Year of adoption	Post
Alabama	10	1987	22		-			-	
Alaska	6	1983	26	9	1986	23	6	1983	26
Arizona	6	1983	26	8	1985	24	8	1985	24
Arkansas	7	1984	25	3	1980	29		-	
Colorado	6	1983	26	8	1985	24		-	
Connecticut		-		3	1980	29	3	1980	29
Delaware		-		15	1992	17	15	1992	17
Florida		-			-			-	
Georgia		-			-			-	
Hawaii	9	1986	23	5	1982			-	
Idaho	0	1977	32	0	1977		12	1989	
Illinois	7	1974	35	1	1978	31		-	
Indiana	10	1987	22	0	1973	36		-	
Iowa	10	1987	22	8	1985	24		-	
Kansas	7	1984	22	4	1981	28		-	
Kentucky	6	1983	26	6	1983	26		-	
Louisiana		-			-		12	1998	11
Maine	0	1977	32		-			-	
Maryland	8	1985	24	4	1981	28		-	
Massachusetts	11	1988	21	3	1980	29	0	1977	32
Michigan	3	1980	29	0	1976	33		-	
Minnesota	6	1983	26	9	1986	23		-	
Mississippi	15	1992	17	10	1987	22		-	
Missouri	6	1983-88	5	8	1985	24		-	
Montana	10	1987	22	3	1980	29	5	1982	27
Nebraska	6	1983	26	10	1987	22		-	
Nevada	6	1983	26	7	1984	25	10	1987	22
New Hampshire	11	1988	21	0	1974	35	0	1974-80	3
New Jersey	8	1985	24	3	1980	29		-	
New Mexico	3	1980	29	6	1983	26		-	
New York	5	1982	27		-			-	
North Carolina		-		8	1985	24		-	
North Dakota	7	1984	25	10	1987	21		-	
Ohio	5	1982	27						
Oklahoma	0	1976	33	13	1990	19		-	
Oregon	1	1978	31	12	1989	20	8	1985-89	4
Pennsylvania		-		0	1974	35		-	
South Carolina	10	1987	22	8	1985	24		-	
South Dakota	6	1983	26	11	1988	21		-	
Tennessee	4	1981	28	7	1984	25		-	
Texas	8	1985	24	7	1984	25		-	
Utah	9	1986	23	12	1989	20		-	
Vermont	8	1985	24	9	1986	23		-	
Virginia	6	1983	26	8	1985	24		-	
Washington	0	1977	32	7	1984	25		-	
West Virginia	9	1986	23	1	1978	31		-	
Wisconsin	8	1985	21	3	1980	29		-	
Wyoming	8	1985	24	12	1989	20	17	1994	15

5 Concluding Remarks

This dissertation examines two linked domains that shape lifetime human capital and inequality: (i) the micro-level social environments in which pupils learn: how teams are formed, who students choose to work with, and how those choices affect learning and well-being; and (ii) the institutional rules of the labor market that determine how the observable signals of schooling are priced and who benefits from them. Chapters 2 and 3 present large-scale lab-in-the-field evidence on team formation and partner effects among schoolchildren in the Barcelona metropolitan area; Chapter 4 uses staggered state-level legal variation in the United States to study how changes in dismissal costs reprice formal schooling and redistribute labor market opportunities. Together, the three chapters map a plausible life-cycle channel: early social sorting inside classrooms shapes exposure to skills and social experiences, while labor-market institutions determine the market value of that.

Chapter 2 uses experimental data from across 19 schools and 1,735 students to delve into how team-formation rules shape who learns with whom in the classroom. Our research shows that when students choose their own partners they tend to select same-gender and higher-performing classmates and systematically exclude socially marginalized pupils (those who experience bullying or are of migrant background), whereas teachers tend to form more ability-heterogeneous and inclusive groups. These compositional differences matter: they change exposure to peer tutoring, social support, and collaborative learning opportunities. While student choice boosts social comfort for some, it risks concentrating advantages and leaving vulnerable students isolated.

Chapter 3 examines the causal effects of partners on both test performance and pupil well-being using randomized pairings in the group-testing phase of the same experimental session as in the previous chapter. We find asymmetric ability effects: lower-ability students typically gain from pairwork, whereas many higher-ability students lose ground in groups. Social ties introduce a welfare trade-off: friendship raises self-reported comfort but is associated with modest declines in measured group performance, while reciprocal enmity harms both comfort and outcomes. These results highlight that group allocation can simultaneously promote socio-emotional benefits and reduce short-run cognitive efficiency.

Chapter 4 investigates how labor-market institutions re-price schooling by exploiting staggered state-level adoption of wrongful-discharge exceptions in the United States. The evidence indicates that higher dismissal costs increase the wage premium for formal schooling, especially for cohorts exposed when young, without producing detectable increases in educational attainment. This pattern is consistent with employers relying more on observable credentials when post-hire adjustment is costly, thereby amplifying returns for the already-educated and producing distributional effects that may disadvantage the least-credentialed.

Concluding Remarks

The research presented in this dissertation makes a substantive contribution to the literature of education economics and its intersection with labor economics, networks and peer effects. By combining large-scale lab-in-the-field evidence on how pupils form and perform in teams with population-level legal variation that re-prices schooling, the thesis links micro-level peer processes to macro-level sorting and returns. This approach not only deepens our understanding of how classroom social organization and labor-market institutions jointly shape skill formation and inequality, but also contributes to a richer comprehension of the intricate relationship between classroom organization, institutional design, and distributional outcomes.

The second chapter of this study builds upon pivotal studies team-formation ([Sacerdote \(2001\)](#), [Hoxby \(2000\)](#), [Baccara and Yariv \(2013\)](#), [Aparicio Fenoll and Zaccagni \(2022\)](#)). These works have established that peers matter for cognitive and non-cognitive outcomes and that endogenous sorting (choice) can amplify homophily by gender, ability or ethnicity; related work on group decision processes highlights how familiarity and repeated interaction alter incentives inside teams ([Thompson et al. \(1998\)](#), [Leaper \(1998\)](#)). This chapter extends these strands by bringing together: (i) direct measures of student preferences (friend/enemy lists and stated desired/undesired partners), (ii) baseline cognitive measures, and (iii) teacher hypothetical groupings; producing a detailed map of the divergence between student self-selection and teacher assignment. By showing how institutional assignment rules (teacher vs. student choice) mechanically reshape exposure to peer-tutoring and inclusive practices, the chapter provides concrete microfoundations for policy debates on whether to allow student choice, mandate teacher assignment, or adopt hybrid algorithms for grouping. The third chapter of this analysis follows on the previous one, contributing to the literature on peer-effects and social incentives.

Prior research (including studies by [Goldman \(1965\)](#), [Webb \(1982\)](#), [Bandiera et al. \(2010\)](#), [Jehn and Shah \(1997\)](#), and [Weidmann and Deming \(2021\)](#)) has examined how group composition and within-team social processes, such as communication, conflict, and social incentives, influence team functioning, effort provision, and information sharing. While these works establish that dynamics inside teams shape productivity and the distribution of contributions, they typically do not jointly estimate short-run cognitive outputs and pupils' socio-emotional well-being in real classroom environments. Chapter 3 fills this gap by identifying causal partner effects on both measured test performance and self-reported comfort. This approach not only broadens our understanding of how within-team social processes shape both cognitive and socio-emotional outcomes in classroom settings, but also offers practical guidance for educators and policymakers designing grouping protocols.

Chapter 4 of this thesis makes a noteworthy contribution to the labor-economics literature by providing a detailed analysis of the interplay between dismissal-cost policies and the returns to education, and by assessing the downstream effect on education decisions. The relevant literature spans classic work on schooling and human capital accumulation ([Becker \(1964\)](#); [Mincer \(1974\)](#)), signaling theory ([Spence \(1973\)](#)), and empirical studies of employment protection and labor-market outcomes ([Autor et al. \(2006\)](#); [Cabrales et al. \(2017\)](#); [Macleod and Nakavachara \(2007\)](#); [Dertouzos and Karoly \(1993\)](#)). The core theoretical prediction, that costly post-hire adjustment increases firms' reliance on observable pre-hire signals, follows directly from signaling logic, yet empirical evidence that links legal changes in dismissal costs to the pricing of schooling and to subsequent schooling responses is relatively scarce. By building upon and synthesizing these concepts, Chapter 4 fills this gap and enriches our understanding of how labor-market institutions re-price observable credentials and thereby influence the distribution of opportunities between high- and low-skilled workers.

To conclude, I want to summarize three main lessons that can be drawn from this dissertation. First, the way in which classroom groups are formed has profound implications for equity in learning opportunities. Student self-selection may lead to homophily by gender and ability and often exclude marginalized pupils, whereas teacher-designed groups are more heterogeneous and inclusive. This study underscores the importance of considering group-formation rules as a policy lever inside schools. Structured teacher interventions, or hybrid approaches that balance student preference with inclusive objectives, can ensure that all pupils have access to productive peer interactions and mitigate the risk of reinforcing existing inequalities.

Second, a central theme emerging from this research is the trade-off between cognitive performance and socio-emotional well-being in collaborative learning. Friendships enhance comfort and morale but may reduce short-run test outcomes, while enmities harm both. It is imperative to account for this dual dimension when designing teamwork exercises. Encouraging collaborative methods that preserve the social benefits of friendship while minimizing off-task behavior (for example through rotation, teacher monitoring, or structured peer-instruction protocols) can help schools achieve a better balance between academic efficiency and student comfort.

Third and finally, labor-market institutions outside the classroom play a decisive role in shaping the returns to schooling. The analysis shows that higher dismissal costs increase the wage premium to formal education but do not automatically generate higher attainment, thereby reallocating opportunities in favor of workers with higher levels of formal education. This finding highlights the need for complementary education and training policies that expand access to schooling and skills development for disadvantaged groups. Such measures are essential to ensure that institutional changes in the labor market translate into more equitable outcomes, rather than reinforcing existing disparities in education and employment.

Bibliography

- Aparicio Fenoll, A. and Zaccagni, S. (2022). Gender mix and team performance: Differences between exogenously and endogenously formed teams. *Labour Economics*, 79(C).
- Arvan, M., Pindek, S., Andel, S., and Spector, P. (2019). Too good for your job? disentangling the relationships between objective overqualification, perceived overqualification, and job dissatisfaction. *Journal of Vocational Behavior*, 115:103323.
- Autor, D. H., Katz, L. F., and Kearney, M. S. (2006). The polarization of the u.s. labor market. *American Economic Review*, 96(2):189–194.
- Baccara, M. and Yariv, L. (2013). Homophily in peer groups. *American Economic Journal: Microeconomics*, 5(3):69–96.
- Bandiera, O., Barankay, I., and Rasul, I. (2010). Social incentives in the workplace. *The review of economic studies*, 77(2):417–458.
- Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, First Edition*. Number beck-5 in NBER Books. National Bureau of Economic Research, Inc.
- Brynin, M. (2002). Overqualification in employment. *Work, Employment and Society*, 16(4):637–654.
- Cabrales, A., Dolado, J. J., and Mora, R. (2017). Dual employment protection and (lack of) on-the-job training: PIAAC evidence for Spain and other European countries. *SERIEs: Journal of the Spanish Economic Association*, 8(4):345–371.
- Card, D. (1999). The causal effect of education on earnings. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3 of *Handbook of Labor Economics*, chapter 30, pages 1801–1863. Elsevier.
- Correll, S. (2001). Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology - AMER J SOCIOL*, 106:1691–1730.
- de Matos Fernandes, C. A., Hoffman, M., and Brouwer, J. (2024). Antecedents of student team formation in higher education. *Learning and Instruction*, 92:101931.
- Dertouzos, J. N. and Karoly, L. A. (1993). Employment effects of worker protection: Evidence from the united states. In Buechtemann, C. F., editor, *Employment Security and Labor Market Behavior: Interdisciplinary Approaches and International Evidence*, pages 217–227. ILR Press, Ithaca, NY.
- Else-Quest, N., Hyde, J., and Linn, M. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological bulletin*, 136:103–27.

Bibliography

- Goldman, M. (1965). A comparison of individual and group performance for varying combinations of initial ability. *Journal of Personality and social Psychology*, 1(3):210.
- Hill, G. W. (1982). Group versus individual performance: Are $n+1$ heads better than one? *Psychological bulletin*, 91(3):517.
- Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. Working Paper 7867, National Bureau of Economic Research.
- Hsieh, C.-J., Fifić, M., and Yang, C.-T. (2020). A new measure of group decision-making efficiency. *Cognitive research: principles and implications*, 5:1–23.
- Izquierdo-Nofuentes, À. (2025). (in progress). The cost of friendship: Who is a better working partner?
- Izquierdo-Nofuentes, À., Arenas, A., and Montolio, D. (2025). (in progress). Building your team: The students dilemma when choosing companion.
- Jehn, K. A. and Shah, P. P. (1997). Interpersonal relationships and task performance: An examination of mediation processes in friendship and acquaintance groups. *Journal of personality and social psychology*, 72(4):775.
- Kugler, A. D. and Saint-Paul, G. (2004). How Do Firing Costs Affect Worker Flows in a World with Adverse Selection? *Journal of Labor Economics*, 22(3):553–584.
- Leaper, C. (1998). Decision-making processes between friends: Speaker and partner gender effects. *Sex Roles*, 39(1-2):125–133.
- Macleod, W. and Nakavachara, V. (2007). Can wrongful discharge law enhance employment? *Economic Journal*, 117(521):F218–F278.
- Mincer, J. A. (1974). *Schooling, Experience, and Earnings*. Number minc74-1 in NBER Books. National Bureau of Economic Research, Inc.
- Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*, 96(1):152–175.
- Ortiz-Ospina, E. and Roser, M. (2016). Global education.
- Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics*, 116(2):681–704.
- Saint-Paul, G. (1997). *Dual Labor Markets: A Macroeconomic Perspective*. MIT Press, Cambridge, MA.
- Saleh, M., Lazonder, A., and De Jong, T. (2005). Effects of within-class ability grouping on social interaction, achievement, and motivation. *Instructional Science*, 33(2):105 – 119.
- Shah, P. P. and Jehn, K. A. (1993). Do friends perform better than acquaintances? the interaction of friendship, conflict, and task. *Group decision and negotiation*, 2:149–165.
- Snyder, T. D., editor (1993). *120 Years of American Education: A Statistical Portrait*. U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, Washington, D.C.

- Spence, M. (1973). Job market signaling*. *The Quarterly Journal of Economics*, 87(3):355–374.
- Thompson, L., Kray, L. J., and Lind, E. A. (1998). Cohesion and respect: An examination of group decision making in social and escalation dilemmas. *Journal of Experimental Social Psychology*, 34(3):289–311.
- Tudge, J. R. H. (1992). Processes and consequences of peer collaboration: A vygotskian analysis. *Child Development*, 63(6):1364–1379.
- Vasco, M., Alfonso, A., Arenas, A., Cabrales, A., Cuesta, J. A., Espín, A. M., Espinosa, M. P., García-Munoz, T., Izquierdo Nofuentes, A., Lomas, P., Jorrat, D., Kovářík, J., Martínez-Estudillo, A. C., Montolio, D., Ramos-Sosa, M. d. P., Roman, P., Sánchez, A., Vázquez-De Francisco, M. J., and Branas-Garza, P. (2025). Teenslab dataset: Economic preferences and cognitive abilities among teenagers. *Scientific Data*. In RR.
- Wang, Z., Kuang, Y., Tang, H., Gao, C., Chen, A., and Chan, K. (2017). Are decisions made by group representatives more risk averse? the effect of sense of responsibility. *Journal of Behavioral Decision Making*, 31.
- Webb, N. M. (1982). Group composition, group interaction, and achievement in cooperative small groups. *Journal of educational psychology*, 74(4):475–484.
- Weidmann, B. and Deming, D. J. (2021). Team players: How social skills improve team performance. *Econometrica*, 89(6):2637–2657.
- Wetmore, W. R. and Summers, J. (2003). Group decision making: friend or foe? In *IEMC'03 Proceedings. Managing Technologically Driven Organizations: The Human Side of Innovation and Change*, pages 405–409. IEEE.
- Zander, L., Höhne, E., Harms, S., Pfof, M., and Hornsey, M. J. (2020). When grades are high but self-efficacy is low: Unpacking the confidence gap between girls and boys in mathematics. *Frontiers in Psychology*, Volume 11 - 2020.

Essays on Education Economics

Abstract

Aquesta tesi explora com les interaccions entre iguals a les escoles i les institucions del mercat laboral configuren conjuntament els resultats educatius i els rendiments de l'escolarització. Els dos primers capítols utilitzen un experiment de camp a gran escala en escoles de l'Àrea Metropolitana de Barcelona per examinar com els estudiants i els professors decideixen formar grups a classe, i com les característiques dels companys influeixen tant en el rendiment acadèmic com en el benestar subjectiu. Els resultats documenten un intercanvi sistemàtic entre benestar i eficiència: els estudiants prefereixen treballar amb amics i amb companys del mateix gènere, i fer-ho augmenta el seu nivell de benestar. No obstant això, aquestes agrupacions poden reduir el rendiment en comparació amb equips més heterogenis, i els grups formats pels professors són més inclusius, ja que milloren la integració d'alumnes marginats. El tercer capítol se centra en el mercat laboral, i fa ús de reformes judicials esglaonades en diversos estats dels EUA que van incrementar els costos d'acomiadament per a les empreses. Els resultats mostren que una protecció laboral més forta augmenta la prima salarial de l'educació formal, especialment entre les cohorts més joves, però no genera increments detectables en el nivell d'escolarització. En conjunt, els capítols enllacen les dinàmiques a l'aula amb les forces institucionals del mercat laboral, ja que ressalten com les preferències socials i les normes institucionals redistribueixen tant les oportunitats com les recompenses al llarg del cicle vital.

Keywords— Educació, Formació de grups, Interacció entre iguals, Amistat, Rendiment, Benestar, Mercat laboral, Costos d'acomiadament