Inclusion Policy Lab: Evaluation Results

Autonomous Community of Navarre – Education: Educational Reinforcement Project for Children in Vulnerable Situations

April 2024

Financiado por la Unión Europea NextGenerationEU
MINISTERIO DE INCLUSIÓN, SEGURIDAD SOCIAL Y MIGRACIONES
Plan de Recuperación, Transformación y Resiliencia
Gobierno de Navarra
cemfi
J-PAL

The General Secretariat of Inclusion of the Ministry of Inclusion, Social Security, and Migration has prepared this report within the framework of the Inclusion Policy Lab, as part of the Recovery, Transformation and Resilience Plan (RTRP), with funding from the Next Generation EU funds. As the agency in charge of carrying out the project, the Government of Navarre has collaborated in the preparation of this report. This entity is one of the implementers of the pilot projects and has collaborated with the General Secretariat of Inclusion in the design of the RCT methodology, actively participating in the provision of the necessary information for the design, monitoring, and evaluation of the social inclusion itinerary. Furthermore, their collaboration has been essential to gathering informed consents, ensuring that participants in the itinerary were adequately informed and that their participation was voluntary.

A research team coordinated by the CEMFI (Center for Monetary and Financial Studies) has substantially contributed to this study. Specifically, María Hernández-de-Benito and Teresa Molina-Millán, both professors at the University of Alicante, under the coordination of Mónica Martínez-Bravo (until January 8, 2024) and Samuel Bentolila, professors at CEMFI. The researchers have been actively engaged in all phases of the project, including the adaptation of the initial proposal to the needs of the evaluation through randomized experiments, the evaluation design, the design of measurement instruments, data processing, and the performance of econometric estimations that lead to quantitative results.

The partnership with J-PAL Europe has played a vital role in the efforts of the General Secretariat of Inclusion to improve social inclusion in Spain. Their team has provided technical support and shared international experience, assisting the General Secretariat in the comprehensive evaluation of pilot programs. Throughout this partnership, J-PAL Europe has consistently demonstrated a commitment to fostering evidence-based policy adoption, facilitating the integration of empirical data into strategies that seek to promote inclusion and progress within our society.

This evaluation report has been produced using the data available at the time of its writing and is based on the knowledge acquired about the project up to that date. The researchers reserve the right to clarify, modify, or delve into the results presented in this report in future publications. These potential variations could be based on the availability of additional data, advances in evaluation methodologies, or the emergence of new information related to the project that may influence the interpretation of the results. The researcher is committed to continue exploring and providing more accurate and updated results for the benefit of the scientific community and society in general.

Index

Executive Summary

- The **Minimum Income Scheme**, established in May 2020, is a minimum income policy that aims to ensure a minimum income to vulnerable groups and provide ways to promote their social and labor integration.
- Within the framework of this policy, the Ministry of Inclusion, Social Security and Migration (MISSM) fosters a strategy to promote inclusion through pilot projects of social innovation, which is conducted in the **Inclusion Policy Lab**. These projects are evaluated according to the standards of scientific rigor and using the methodology of Randomized Controlled Trials.
- This document presents the evaluation results and main findings of the "Educational Reinforcement Project for Children in Vulnerable Situations", carried out in **cooperation between the MISSM and the Government of Navarre**.
- This study evaluates the effectiveness of an **intensive online educational reinforcement program through tutoring** in groups of 2 students, comparing it with tutoring groups of 3 students, and with a group of 2 students whose tutor receives additional training in socio-emotional aspects. The participating students received tutoring for eight weeks in the different formats mentioned. The effect of these interventions has been compared with a control group that did not receive any intervention.
- The project was conducted in the **Autonomous Community of Navarre**, with **1,344 students** from **46 public schools** located in different areas of Navarre. The program was carried out in two waves, with 545 students participating in the first wave and 799 in the second.
- The participating students were between 5th year of primary school and 2nd year of secondary school (ESO). Of these, 66% of students in the first wave were part of the census of students with Specific Educational Support Needs, and 59% of the second wave. In the standardized mathematics test carried out prior to the intervention, the average score in both waves was 3 out of 10, evidence the low level of participating students on the subject.
- The attrition rate was nearly 6% for students who did not take the post-intervention level test, and around 40% for students who did not complete the final survey. The attrition was higher in the control group of the second wave of the program.
- The main results of the evaluation are as follows:
    - **Improvement in mathematical results**: In some experimental groups, it is observed that **tutoring** has positive and statistically significant effects on the **mathematics tests results**, as well as on overall school grades. In the second wave, **tutoring with 3 students** is also found to be effective in improving the **overall average school grades**. There is no negative effect on **school performance** of tutoring in groups of 3 students, compared to the smaller group of 2 students. There are also no significant differences in school performance for students tutored by teachers trained in socio-emotional skills.
    - **Increased confidence in math skills and well-being at school**: In the first wave, tutoring with 3 students improved **students' confidence** in their mathematical performance and reduced their perception of school-related stress. In the second

wave, the **socio-emotional component in tutor training** had greater effects on students' confidence in their mathematical abilities, as well as a greater appreciation for the subject.

- **Academic aspirations**: The tutoring programs had no impact on academic aspirations, as measured by expectations of completing high school (desire to attend baccalaureate) or desire to attend university.

1 Introduction

General Regulatory Framework

The Minimum Income Scheme (MIS), regulated by Law 19/2021[1], is an economic benefit whose main objective is to prevent the risk of poverty and social exclusion of people in situations of economic vulnerability. Thus, it is part of the protective action of the Social Security system in its non-contributory modality and follows the recommendations of various international organizations to address the problem of inequality and poverty in Spain.

The provision of the MIS has a double objective: to provide economic support to those who need it most and to promote social inclusion and employability in the labor market. This is one of the social inclusion policies designed by the General State Administration, together with the support of the Autonomous Communities, the Third Sector organizations, and local corporations[2]. It is a central policy of the Welfare State that aims to provide minimum economic resources to all individuals in Spain, regardless of where they live.

Within the framework of the National Recovery, Transformation, and Resilience Plan (RTRP),[3] the General Secretariat of Inclusion (SGI) of the Ministry of Inclusion, Social Security and Migration (MISSM) participates significantly in Component 23 "New public policies for a dynamic, resilient and inclusive labor market", framed in Policy Area VIII: "New care economy and employment policies".

Investment 7 "Promotion of Inclusive Growth by linking socio-labor inclusion policies to the Minimum Income Scheme" is among the reforms and investments proposed in this Component 23. Investment 7 promotes the implementation of a new model of inclusion based on the MIS which reduces income inequality and poverty rates. Therefore, the MIS goes beyond being a mere economic benefit and supports the development of a series of complementary programs that promote socio-labor inclusion. However, the range of possible inclusion programs is very wide, and the government decides to pilot different programs and interventions to evaluate them and generate knowledge that allows prioritizing certain actions. With the support of investment 7 under component 23, the MISSM establishes a new framework for pilot inclusion projects constituted in two phases through two royal decrees covering a set of pilot projects based on experimentation and evaluation:

[1] Law 19/2021, dated December 20, establishing the Minimum Income Scheme (BOE-A-2021-21007).

[2] Article 31.1 of Law 19/2021, of December 20, 2021, establishing the Minimum Income Scheme.

[3] The Recovery, Transformation, and Resilience Plan refers to the Recovery Plan for Europe, which was designed by the European Union in response to the economic and social crisis triggered by the COVID-19 pandemic. This plan, also known as Next Generation EU, sets out a framework for the allocation of recovery funds and for boosting the transformation and resilience of member countries' economies.

- **Phase I: Royal Decree 938/2021**[4], through which the MISSM grants subsidies for the execution of 16 pilot projects of inclusion pathways corresponding to autonomous communities, local organizations, and the Third Sector of Social Action organizations. This royal decree contributed to the fulfillment of milestone number 350[5] and monitoring indicator 351.1[6] of the RTRP.
- **Phase II: Royal Decree 378/2022**[7], which grants subsidies for a total of 18 pilot projects of inclusion pathways executed by autonomous communities, local organizations, and the Third Sector of Social Action organizations. Along with the preceding Royal Decree, this one helped the RTRP's monitoring indicator number 351.1 to be fulfilled.

To support the implementation of evidence-based public and social policies, the Government of Spain decided to evaluate the social inclusion pilot projects using the Randomized Controlled Trial (RCT) methodology. This methodology, which has gained relevance in recent years, represents one of the most rigorous tools to measure the causal impact of a public policy intervention or a social program on indicators of interest, such as social and labor insertion or the well-being of beneficiaries.

Specifically, RCT is an experimental method of impact evaluation in which a representative sample of the population potentially benefiting from a public program or policy is randomly assigned either to a group receiving the intervention or to a comparison group that does not receive the intervention for the duration of the evaluation. Thanks to the randomization in the allocation of the program, this methodology can statistically identify the causal impact of an intervention on a series of variables of interest. This methodology enables us to analyze the effect of this measure, which helps to determine if the policy is adequate to achieve the planned public policy objectives. Experimental evaluations enable us to obtain rigorous results of the intervention effect, i.e., what changes the participants have experienced in their lives due to the intervention. In addition, these evaluations provide an exhaustive analysis of the program and its effects, providing insights into why the program was effective, who

[4] Royal Decree 938/2021, of October 26, 2021, which regulates the direct granting of subsidies from the Ministry of Inclusion, Social Security and Migration in the field of social inclusion, for an amount of 109,787,404 euros, within the framework of the Recovery, Transformation and Resilience Plan (BOE-A-2021-17464).

[5] Milestone 350 of the PRTR: "Improve the rate of access to the Minimum Income Scheme and increase the effectiveness of the IMV through inclusion policies, which, according to its description, will translate into supporting the socio-economic inclusion of the beneficiaries of the IMV through itineraries: eight collaboration agreements signed with subnational public administrations, social partners and Third Sector organizations of Social Action to carry out the itineraries. The objectives of these partnership agreements are: i) to improve the MVI access rate; ii) increase the effectiveness of the MVI through inclusion policies."

[6] Monitoring indicator 351.1 of the RTRP: "at least 10 additional collaboration agreements signed with subnational public administrations, social partners and Third Sector organizations to carry out pilot projects to support the socio-economic inclusion of the beneficiaries of the MVI through itineraries".

[7] Royal Decree 378/2022, of May 17, 2022, regulating the direct granting of subsidies from the Ministry of Inclusion, Social Security and Migration in the field of social inclusion, for an amount of 102,036,066 euros, within the framework of the Recovery, Transformation and Resilience Plan (BOE-A-2022-8124).

has benefited most from the interventions, whether it has indirect or unexpected effects, and which components of the intervention work and which do not.

These evaluations have focused on the promotion of social and labor inclusion among MIS beneficiaries, recipients of regional minimum incomes and other vulnerable groups. In this way, the MISSM establishes a design and impact evaluation of results-oriented inclusion policies, which offers evidence for decision-making and its potential application in the rest of the territories. The promotion and coordination of 32 pilot projects by the Government of Spain has led to the establishment of a laboratory for innovation in public policies of global reference named as the Inclusion Policy Lab.

For the implementation and development of the Inclusion Policy Lab, the General Secretariat of Inclusion has established a governance framework that has made it possible to establish a clear and potentially scalable methodology for the design of future evaluations, and the promotion of decision-making based on empirical evidence. The General State Administration has had a triple role as promoter, evaluator, and executive of the different programs. Different regional and local administrations and the Third Sector organizations have implemented the programs, collaborating closely in all their facets, including evaluation and monitoring. In addition, the Ministry has had the academic and scientific support of the Abdul Latif Jameel Poverty Action Lab (J-PAL) Europe and the Centre for Monetary and Financial Studies (CEMFI), as strategic partners to ensure scientific rigor in the assessments. Likewise, the Inclusion Policy Lab has an Ethics Committee[8], which has ensured the strictest compliance with the protection of the rights of the people participating in the social inclusion pathways.

This report refers to the "Educational Reinforcement Project for Children in Vulnerable Situations", implemented within the framework of Royal Decree 378/2022[9] by the Autonomous Community of Navarre. In this sense, the Department of Education and the Department of Social Rights of the regional government have been responsible for the implementation of the project. This report contributes to the fulfillment of milestone 351 of the RTRP "Following the completion of at least 18 pilot projects, the publication of an evaluation on the coverage, effectiveness and success of the MIS, including recommendations to increase the level of application and improve the effectiveness of social inclusion policies".

[8] Regulated by Order ISM/208/2022, of March 10, 2022, which creates the Ethics Committee linked to social inclusion itineraries, on 20/05/2022 it issued a favorable report for the realization of the project that is the subject of the report.

[9] On December 16, 2022, an Agreement was signed between the General State Administration, through the SGI, and the Autonomous Community of Navarre for the implementation of a project for social inclusion within the framework of the Recovery, Transformation and Resilience Plan, which was published in the "Official State Gazette" on December 26, 2022 (BOE no. 309).

Project context

Education stands as a fundamental human right and a foundation upon which societies are constructed. Effective and high-quality education is recognized as a central tool for socio-economic development, serving as a potent means to alleviate poverty, enhance health, and foster equality. In line with these principles, the education system, training programs, and teaching personnel emerge as pivotal agents not solely in individual development, but also in the pursuit of social inclusion.

In the current context, characterized by a growing student population with diverse needs[10], it's imperative to rethink and adapt teacher training and teaching approaches to ensure the provision of effective, high-quality education. This becomes especially relevant in the fight against inequalities, particularly in addressing the educational disadvantages experienced by students living in vulnerable environments.

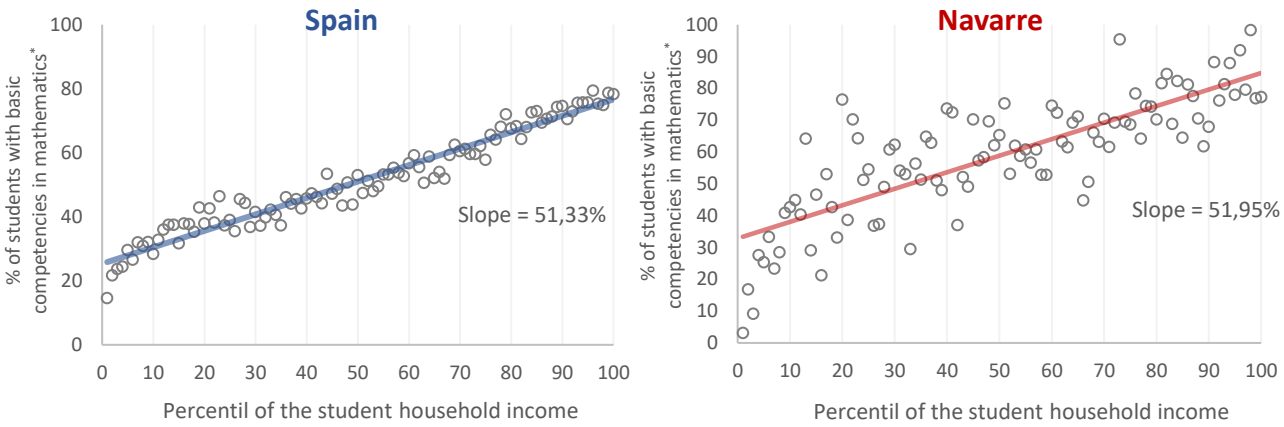
Numerous studies[11] have analyzed the relationship between the students' socio-economic status and their educational attained levels. The consensus suggests that academic progress is highly influenced by the socio-economic circumstances in which students stand. In this line, the recent study from COTEC (2023) on social mobility and inequality of opportunities in Spain, has revealed that 26% of the variance in opportunities can be attributed to factors outside of the student's influence, such as household income.

As illustrated in the following figure, both in Spain and in Navarre, the mathematical competencies of students are closely linked to household income. Students from wealthier households (above the median or 50th percentile) are approximately 51% more likely to have basic math proficiency compared to those from poorer households (below the median). Although the slope of this relationship is quite similar in both Spain and in the region of Navarre, the later stands out for having a greater dispersion of observations as well as a higher average basic knowledge of mathematics, of 8 percentage points (p.p.) compared to the Spanish average.

[10] The rise in diversity within the student community is a phenomenon addressed and explored in documents such as "The Promotion of Diversity and Inclusion in Schools in Europe" (available at https://data.europa.eu/doi/10.2797/786022) or in the "Strategic Plan: Attention to Diversity in Navarre (available at https://gobiernoabierto.navarra.es/sites/default/files/plan_estrategico_de_atencion_a_la_diversidad.pdf)

[11] Out of many studies, some that stand out are: Roemer, J. (2000). Equality of Opportunity. In the Meritocracy and economic inequality. Princeton University Press. Roemer, J. E. (2002). Equality of opportunity: A progress report. Social Choice and Welfare, 455-471. Roemer, J. E., & Trannoy, A. (2016). Equality of opportunity: Theory and measurement. Journal of Economic literature, 54(4), 1288-1332. Penguin UK. Sen, A. (2000). Merit and Justice. In the Meritocracy and economic inequality. Princeton University Press. Soria-Espin, J. (2022). Intergenerational Mobility, Gender Differences, and the Role of Out-Migration: New Evidence from Spain. Zamarro, G., Hitt, C., & Mendez, I. (2019). When students don't care: Reexamining international differences in achievement and student effort. Journal of Human Capital, 13(4), 519-552.

Figure 1: Relationship between household income and basic skills in mathematics



Grades: (*) Basic competencies in mathematics measured as students with PL3 level or higher in the subject. Results obtained from the results of PISA 2018. Source: COTEC (2023)

To address this socio-educational gap, small group tutoring has emerged as a powerful tool to mitigate disparities in academic performance and promote inclusion and social mobility. Unlike traditional teaching methods, where students often assume passive roles, tutoring emphasizes cooperation, collaboration, and continuous encouragement among group members. This approach enables students to reach higher levels of academic competencies by actively engaging them in the learning process.

Regulatory framework associated with the project and governance structure

Within the European Union, member states operate autonomously concerning educational policies and initiatives. However, the EU facilitates cooperation to ensure coherence among countries. In February 2021, the European Commission published the 2021-2030 Strategy for Education and Training, outlining general objectives for member states. Specifically, in relation to the theme of this report, the European strategy aims to reduce the school dropout rate to below 9% and decrease the percentage of students with low performance in reading, mathematics, and science to 15%.

Likewise, a few months after, the European Commission published Recommendation (EU) 2021/1004 in June 2021[12], establishing a European Child Guarantee. This document encourages member states to implement national plans to ensure access to basic health and education rights for children at risk of poverty and social exclusion.

[12] The European Child Guarantee provides guidance and tools for EU countries to implement strategic plans aimed at ensuring access to essential health and education services for children.
https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32021H1004

In response to the European recommendations, the Government of Spain introduced the National Action Plan for the Implementation of the European Child Guarantee (2022-2030) in July 2022.[13] Among the challenges identified in this plan, a primary focus is to "promote educational equity through comprehensive and flexible education, capable of adapting to individualized needs, especially for the most vulnerable children."

On the other hand, the Education Law[14], which lays the regulatory foundations of the Spanish education system and was last updated on December 29, 2020, by Organic Law 3/2020, includes several relevant aspects regarding the program object of this study. Articles 81.2 and 81.3 are particularly noteworthy, as they emphasize the need for socio-educational actions, such as accompaniment and tutoring, in schools, geographical areas, or social environments where there is a concentration of students subject to socio-educational vulnerability.

From a regional perspective, Provincial Decree 66/2010 of the Autonomous Community of Navarre, issued on October 29, is particularly noteworthy. This decree establishes regulations for educational and vocational guidance in the region's schools, highlighting tutoring and specialized advice as fundamental principles of educational guidance.

Finally, all European, national, and regional regulations are in line with the framework established in the 2030 Agenda and the Sustainable Development Goals (SDGs).

The pilot project of this report is aligned with European, national, and regional strategies in the field of education, tutoring and social integration of schoolchildren, as well as with the 2030 Agenda for Sustainable Development, specifically contributing to SDGs 1, 4, and 10.

Given the educational context in Spain and Navarre, the needs in this regard and the potential benefits identified of reinforcement through tutoring in small groups, the Autonomous Community of Navarre has led a project aimed at developing and testing the effectiveness of an online tutoring model for students from vulnerable backgrounds. The project aims to help students bridge the educational gap caused by socioeconomic factors.

The scientific objective of the project is to evaluate the effectiveness and efficiency of the online tutoring model in small groups. In this sense, it seeks to replicate and expand the results obtained in similar programs carried out during the confinement due to COVID-19, as well as to compare the results with those obtained in similar programs applied in face-to-face mode. In addition, it is intended

[13] The State Action Plan for the Implementation of the European Child Guarantee (2022-2030) represents the main tool with which Spain implements Recommendation (EU) 2021/1004 establishing a European Child Guarantee aimed at breaking the cycle of child poverty.
https://www.mdsocialesa2030.gob.es/derechos-sociales/infancia-y-adolescencia/docs/PlanAccion_MAS.pdf.

[14] Organic Law 2/2006 of 3 May 2006 on Education is a state organic law that regulated educational teaching in the different age groups of education. This law has undergone two amendments to date, through laws published on December 10, 2013 and December 30, 2020.
https://www.boe.es/eli/es/lo/2006/05/03/2/con.

to promote the transfer of knowledge to the process of public policy development and to be accountable for the results of the project.

The governance framework stablished for the correct execution and evaluation of the project includes the following actors:

- The **Autonomous Community of Navarre** is responsible for executing the project through the Department of Education and the Department of Social Rights of the regional government. Due to their specific competencies, these departments are closely aligned with the project's theme and objectives. The **Department of Education** oversees non-university education, carrying out tasks in the field of schooling, digitalization, and other educational services. Meanwhile, the **Department of Social Rights** manages social services, family support, child and teenage care, and the promotion of personal autonomy within the region.

- The **Ministry of Inclusion, Social Security and Migration** (MISSM) is the funding source of the project and responsible for the RCT evaluation. For this reason, the General Secretariat for Inclusion assumes the following commitments to the Autonomous Community of Navarre:
    - Provide the beneficiary entity with support for the design of the actions to be carried out, for the execution and monitoring of the object of the subsidy, as well as for the profiling of the potential participants of the pilot project.
    - Design the randomized controlled trial (RCT) methodology of the pilot project in coordination with the beneficiary entity and scientific partners. Also, carry out the evaluation of the project.
    - Ensure strict compliance with ethical considerations by obtaining the approval of the Ethics Committee.

- **CEMFI and J-PAL Europe** are scientific and academic institutions that support MISSM in the design and the RCT evaluation of the project.

In view of the above, the current report follows the following structure. **Section 2** provides a description of the project, detailing the issue to be addressed, the specific interventions associated with each of the employment models implemented, and the target audience to which the intervention is directed. The objective is to present a diagnosis of the problems associated with homelessness that justifies the need to implement and evaluate this intervention. Next, **section 3** contains information related to the **Evaluation Design**, defining the Theory of Change linked to the project and the hypotheses, sources of information and indicators used. **Section 4** describes the **Implementation of the intervention**, analyzing the sample, the results of randomization, and the degree of participation and attrition of the intervention. This section is followed by **section 5**, where **the results of the evaluation** are presented, with a detailed analysis of the econometric analysis conducted and the results for each of the indicators used. Finally, the **Conclusions** of the project evaluation are described in **section 6**. Besides, in the **Economic and regulatory management** appendix additional information is provided regarding the management instruments and governance of the pilot project.

Ethics Committee linked to the Social Inclusion Itineraries

During research involving human subjects in the field of biology or the social sciences, researchers and workers associated with the program often face ethical or moral dilemmas in the development of the project or its implementation. For this reason, in many countries it is common practice to create ethics committees that verify the ethical viability of a project as well as its compliance with current legislation on research involving human beings. The Belmont Report (1979) and its three fundamental ethical principles – respect for individuals, profit, and justice – constitute the most common frame of reference in which ethics committees operate, in addition to the corresponding legislation in each country.

With the aim of protecting the rights of participants in the development of social inclusion itineraries and ensuring that their dignity and respect for their autonomy and privacy are guaranteed, Order ISM/208/2022 dated March 10 creates the Ethics Committee linked to the Social Inclusion Itineraries. The Ethics Committee, attached to the General Secretariat of Inclusion and Social Welfare Objectives and Policies, is composed of a president – with an outstanding professional career in defense of ethical values, a social scientific profile of recognized prestige and experience in evaluation processes – and two experts appointed as members.

The Ethics Committee has conducted analysis and advice on the ethical issues that have arisen in the execution, development, and evaluation of the itineraries, formulated proposals in those cases that present conflicts of values and approved the evaluation plans of all the itineraries. In particular, the Ethics Committee issued its approval for the development of this evaluation on September 8, 2023.

2 Description of the program and its context

This section describes the program that the Autonomous Community of Navarre implemented in the framework of the pilot project. Furthermore, it describes the target population and the territorial framework and provides a detailed description of the intervention.

2.1 Introduction

The objective of this project is to assess the effectiveness of an intensive online educational reinforcement program, implemented by teaching professionals and with a focus on mathematics. Targeted towards students in 5th and 6th grades of primary school, as well as 1st and 2nd years of secondary school (ESO), the program aims to support those from vulnerable backgrounds.

Although face-to-face tutoring programs, whether in individual sessions or small groups, have been proven to have a very positive impact on the learning process, the evidence regarding online tutoring

is limited. In this context, this project is an initiative aimed at expanding knowledge regarding the possibilities of improving students' academic results through tutoring, outside the classroom and online. In addition, it is intended to promote the transfer of knowledge to the policymaking process and to learn about the results of the project.

Within the general scope of intervention addressed in this report, numerous empirical studies have been conducted in line with the proposed intervention, utilizing the RCT methodology. Notably, several studies on the effectiveness of face-to-face reinforcement classes stand out, with meta-analyses by Nickow et al. (2020) and Alegre et al. (2019) demonstrating consistent and substantial positive impacts on learning outcomes. These studies recognize tutoring as a versatile, cost-effective, and potentially transformative educational tool in the current educational landscape.

In the national context, several academic programs and publications have investigated the efficacy of face-to-face tutoring in small groups for school-age students. Of particular significance is the "*Leemos en Pareja*" program, which was implemented in schools throughout Spain, including numerous schools in the Autonomous Community of Navarre. This program was evaluated by Zambrano et al. (2013) through Randomized Controlled Trials (RCTs) methodology.

While scientific evidence for online tutoring is more limited, the COVID-19 pandemic has spurred the study of this model of educational reinforcement sessions. Carlana et al. (2021) conducted an experimental program with RCT methodology in Italy, reporting a substantial increase in students' academic performance, as well as improvements in socio-emotional skills, aspirations, and psychological well-being, especially in students of low socioeconomic status and immigrants. Conversely, the experiment by Kraft et al. (2022) in the United States yielded non-significant positive effects on student performance, suggesting that online tutoring may not have as broad an impact as face-to-face tutoring. However, it concluded that with this method broader accessibility could be achieved at a lower cost.

Finally, sharing many similarities with the pilot program outlined in this report, the MENTTORES project conducted in Spain from 2021 to 2022, and evaluated by Gortazar et al. (2023), holds particular significance. The MENTTORES project consisted of an online tutoring program in groups of 2 people, focused on mathematics, and aimed at students from vulnerable backgrounds. The results of the program show a substantial increase in students' academic performance, as well as significant improvements in other social and educational indicators.

Due to its close relation and shared similarities, the MENTTORES program, its evaluation and results sets a starting point for the program under review in this report. The objective of this program is to replicate the findings of MENTTORES and to extend the study to contribute to a deeper understanding of the effects of changes in tutoring design.

Additionally, although not employing the RCT methodology, the publication by Q-SER (2020), where a model to identify training needs and solve them through tutorials is designed, and the article by Betthäuser et al. (2023), on the effects of online training on students are also considered pivotal literature in the context of this pilot project.

2.2 Target population and territorial scope

The target population of the project is students attending public schools located in the Autonomous Community of Navarre and in areas with high rates of vulnerability. The participating students are between the third cycle of primary education and the first cycle of secondary education (from 5th year of primary school to 2nd year of ESO). Specifically, the criteria for the selection of these schools with a vulnerable group are as follows:

- **Poverty rate:** The school must be in one of the areas with a relatively high poverty rate.
- **Social exclusion:** The school must have a negative score on the vulnerability index.

The strategy involves deploying the project in schools characterized by having the highest Vulnerability Indexes and located in the areas with the highest poverty rates. To determine these areas, the following indicators from the Navarre Observatory of Social Reality are considered:

- Relative at-risk-of-poverty rate for children under 16 years of age (%): measured as the percentage of children living in households whose total annual equivalent income is below 60% of the median.
- At-risk-of-severe poverty rate for children under 16 years of age (%): measured as the percentage of people living in households whose total annual equivalent income is below 40% of the median.

The schools identified as potential beneficiaries of the pilot project are those located in areas where the sum of both indicators is 70% or more in the first wave, and 50% or more in the second wave. The reduction in the at-risk-of-poverty rate in the eligibility of second-wave schools is explained by the need to obtain a sufficiently large sample of students for the conclusions of the evaluation to be statistically significant.

In addition to the at-risk-of-poverty rate, the ACNEAE Index was used as an indicator of social exclusion. This index is prepared by the Educational Inspection Service of the Department of Education and published annually by Provincial Decree. The index is constructed such that the more negative the index, the greater the risk of social exclusion.

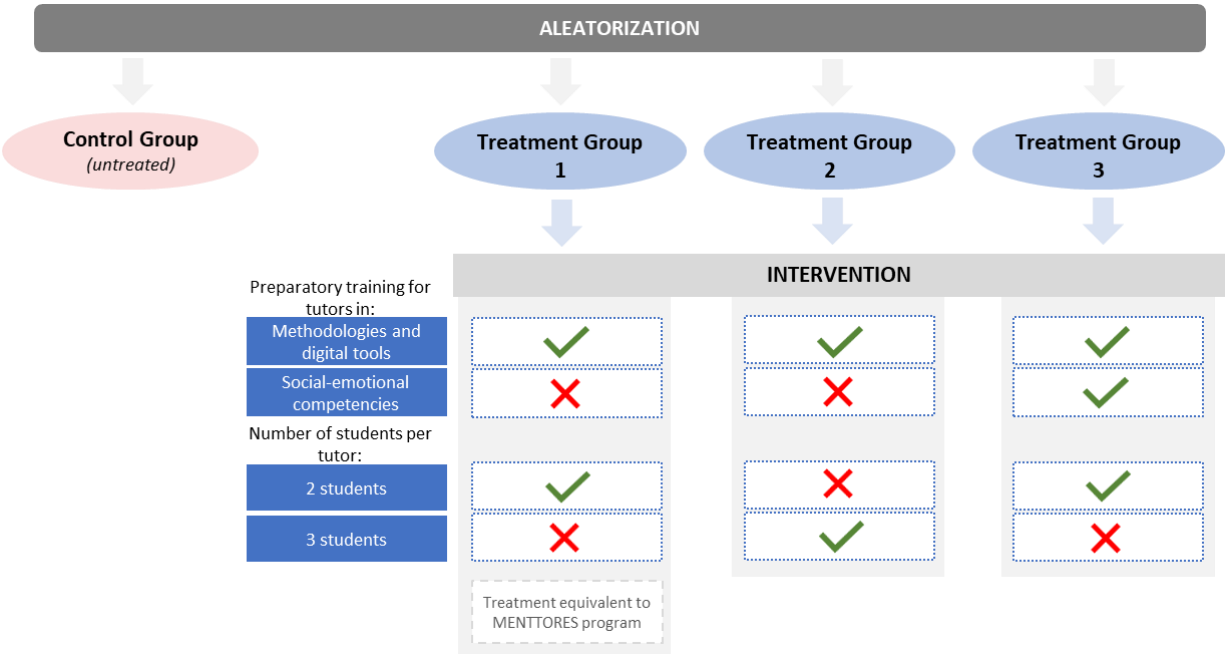
Further details about the students involved in this project and the recruitment process can be found in **section 3.5** as part of the evaluation design.

2.3 Description of interventions

Participants in the project are divided into a control group and three different treatment groups. The control group does not receive any type of intervention. All treatment groups participate in online tutorials given by teaching professionals, during extracurricular hours and over a period of eight weeks. Differences between the treatment groups are based on the number of students per group and the preparatory training received by the teachers. The differences between treatment groups aim to measure their effect on the students' outcomes.

In this way, treatment group 1 receives an intensive program of online mathematics tutoring, in groups of two students per tutor and with tutors who have received a specific training of 15 hours in methodologies and digital tools, with a pedagogical component to teach classes effectively online. This is considered the 'baseline' treatment and replicates the action carried out in the MENTTORES program evaluated by Gortazar et al. (2023). Treatment group 2 differs from the first by setting groups of three students per tutor instead of two. Treatment group 3 is set in groups of two students per tutor but differs from the previous two treatment groups by providing tutors with an additional 12-hour preparatory training focused on social-emotional competencies. These 12 hours are in addition to the 15 hours that the teacher also receives in methodologies and digital tools, thus representing a total of 27 hours of training. **Figure 2** It summarizes the actions carried out according to the experimental group that receives them.

Figure 2: Outline of the intervention



The tutorials last 90 minutes and are provided twice a week. Each tutor handles a maximum of six groups of students, and the tutoring is provided remotely from special rooms enabled by the Government of Navarre. In addition, teachers travel on Fridays to the different schools to coordinate with their students' school tutors.

The project is developed in two different waves or phases. The first wave takes place between March and May 2023 and is carried out with the control group and the first two treatment groups (treatment group 1 and 2). On the other hand, the second wave, which takes place between September and

November 2023, is implemented with all experimental groups (control group and treatment groups 1, 2 and 3).[15]

Additionally, it must be noted that while in the first wave the poverty rate established for the school to be the target of the program is 70%, in the second wave this criterion is reduced, establishing a poverty rate of more than 50%.

In both waves, the intervention phases are as follows:

- The Department of Education contacts eligible schools to inform them about the project and ask for their participation.
- Schools that agree to participate inform the families of students enrolled in the courses listed above. Families who agree to participate sign the informed consent.
- Once the sample of participating students has been defined, the random assignment is carried out and the tutoring groups are defined.
- Tutors are recruited by the Department of Education. Then these receive preparatory training and are coordinated with the schools. At the same time, the schools carry out the pre-assessment test on all the students of the corresponding courses, as well as the baseline survey of the participating students.
- Tutorials are carried out, with an initial first face-to-face meeting between tutors and students to break the digital barrier. Subsequently, the 8 weeks of intervention are carried out.
- After the end of the tutoring period, schools carry out the final assessment test and the final line survey of participating students. In addition, satisfaction surveys are carried out with teachers and schools, and surveys on the intervention are collected from tutors and families of participating students.

3 Evaluation design

This section describes the design of the impact assessment of the projects described in the previous section. The section describes the Theory of Change, which identifies the mechanisms and aspects to measure, the hypotheses to test in the evaluation, the sources of information to build the indicators, and the design of the experiment.

3.1 Theory of Change

This report, with the aim of designing an evaluation that enables us to understand the causal relationship between the intervention and its final objective, develops a Theory of Change. The Theory of Change makes it possible to schematize the relationship between the needs identified in the target

[15] For more information on the time frame of the evaluation, please refer to **section** ¡Error! No se encuentra el origen de la referencia..

population, the benefits, or services that the intervention provides, and the immediate and medium-long term results sought by the intervention, to understand the relationships between them, the assumptions on which they are based, and to outline measures or outcome indicators.

Theory of Change

A Theory of Change begins with the correct identification of the needs or problems to be addressed and their underlying causes. This situational analysis should guide the design of the intervention, i.e., the activities or products that are provided to alleviate or resolve the needs, as well as the processes necessary to properly implement the treatment. Next, we identify the expected effect(s) based on the initial hypothesis, i.e., what changes – in behavior, expectations, or knowledge – are expected to be obtained in the short term with the actions conducted. Finally, the process concludes with the definition of the medium- to long-term results that the intervention aims to achieve. Sometimes, the effects directly obtained with the actions are identified as intermediate results and one identifies the indirect effects in the final results.

The development of a Theory of Change is a fundamental element of impact evaluation. At the design stage, the Theory of Change helps to formulate hypotheses and identify the indicators needed for the measurement of results. Once the results are achieved, the Theory of Change makes it easier, if results are not as expected, to detect which part of the hypothetical causal chain failed, as well as to identify, in case of positive results, the mechanisms through which the program works. Likewise, the identification of the mechanisms that made the expected change possible allows a greater understanding of the possible generalization or not of the results to different contexts.

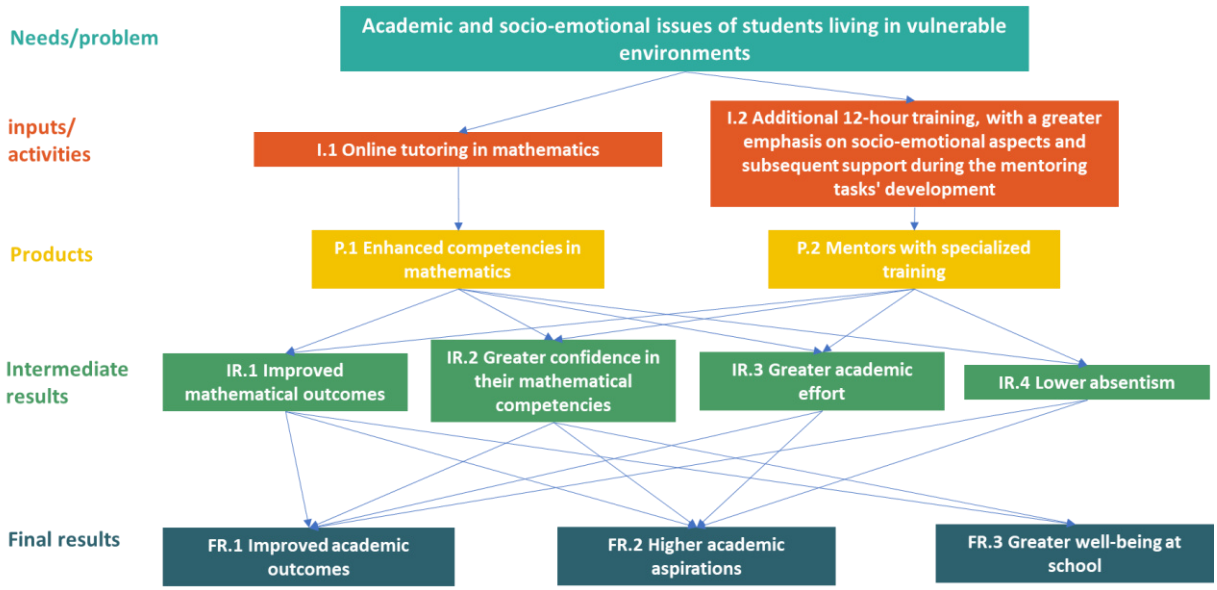
The Theory of Change of this project is based on the identification of the academic and socio-emotional problems that students who reside in vulnerable environments may suffer. To address these issues, the project focuses on providing reinforcement sessions or tutorials, delivered by teaching staff equipped with an understanding of the educational and socio-emotional needs of vulnerable students. Within this framework, a Theory of Change model is devised to align with the different treatments outlined in the pilot project.

Figure 3 illustrates the project's Theory of Change. It starts with the recognition of academic difficulties and socio-emotional needs among vulnerable students, leading to two primary activities: providing online tutorials in mathematics (applicable to all treatment groups) and offering additional tutor training focused on socio-emotional competencies (specific to treatment group 3). These activities are expected to yield two products: reinforced mathematics skills among students and trained tutors equipped to address socio-emotional needs during tutoring sessions.

The entire development of the project leads to a series of intermediate results, acting as precursors of final outcomes. These intermediate results include improved math scores, enhanced confidence in math skills, increased academic effort, and reduced school absenteeism.

Last, the final outcomes include a positive impact on students' overall academic performance (final result 1) and higher academic aspirations (final result 2). Additionally, stemming specifically from the intermediate results of improved mathematical outcomes and increased confidence in mathematical competencies, greater well-being within the school environment is expected. This constitutes the third final result.

Figure 3. Theory of Change



3.2 Hypothesis

The main goal of the project is to enhance the academic experience of students hailing from vulnerable backgrounds. Given the array of challenges these students face, the program delineates its anticipated outcomes into five facets of students' academic lives targeted for improvement through the pilot project.

In each of these areas, a set of primary and secondary hypotheses has been formulated for evaluation. The following are the hypotheses considered:

1. Better math results

The main hypothesis points out that, as a result of tutoring, students achieve an improvement in their math grades. Also, in addition to the main hypothesis, two secondary hypotheses are defined. The first postulates that the intervention has a positive impact on math grades in the long term. The second postulates that students improve their overall academic performance.

2. Increased confidence in their math competencies

Concerning confidence in mathematics, the main hypothesis of this section proposes that the interventions result in an increased self-efficacy of students in mathematics. This hypothesis implies that students' proficiency in problem-solving and analytical skills in mathematics experiences a widespread enhancement.

3. Improved academic performance

The hypothesis proposes that the treatment leads to a decrease in school absenteeism. Complementing the primary hypothesis, a secondary hypothesis is outlined that suggests an increase in the time students allocate to task performance or homework completion.

4. Increased academic aspirations

The primary hypothesis of this section points out that improved academic performance fosters the development of elevated educational expectations. As a secondary hypothesis, it is postulated that students will exhibit an increased desire to pursue higher education at the university level.

5. Greater well-being at school

Finally, concerning well-being at school, the main hypothesis proposes that enhanced academic performance will lead to reduced stress, driving students to gain confidence in their abilities and perception of studies and school. Additionally, as a secondary hypothesis, it is postulated that the perception of support through tutoring will foster greater well-being and satisfaction within the school environment.

3.3 Sources of information

Three data collection methods are utilized in the pilot project. Firstly, administrative data is collected from the official institutions and schools, including socioeconomic information on the family unit, vulnerability status, and academic report cards. Secondly, surveys are conducted with students, their families, and the tutors involved in the project. Finally, tests are administered to students to gather standardized information on their mathematics proficiency.

All the information used to construct outcome indicators and to analyze the project's results is gathered through the mentioned data collection methods, both pre-intervention and post-intervention, as for example with baseline surveys and final-line surveys. This framework remains consistent across both waves of the pilot project. These sources of information are set out and detailed below:

- **Student survey**: aims to collect information on students' perceptions of different socio-educational aspects and is carried out at two points in time – prior to the intervention (baseline survey) and after it (end-line survey) – to observe possible changes in students' responses. The survey is carried out in the school, within the classroom, simultaneously to all participating students (from all treatment groups and the control group). The survey has three

differentiated sections: in the first section personal and academic topics are discussed; the second sections is focused on the relationship of students with mathematics; and the final section includes the topics related the well-being of students and their relationship with the studies and school environment.
In more detail, the first section includes, among others, questions related to socio-emotional skills, ability to perform work (concentration, diligence, interest), dedication to studies and educational aspirations. In the second section, the questions asked include: mathematical abilities, interest in the subject, way of approaching classes and tasks, etc. Finally, the third block of questions addresses issues such as the valuation of relationships and satisfaction with social and family life, and the perception they have of school, the importance they give to it, and other issues that reflect the student's relationship with studies and school.

- **Mathematics tests**: like the surveys, students take a mathematics test before and after the intervention. These placement tests focus on assessing mathematical competencies by solving different calculations or problems. Tailored assessments are devised to match the varying levels of knowledge corresponding to the students' school cycles.
- **Report cards:** the report cards of the participating students are collected, both from the year in which they receive the tutoring and from subsequent and, where appropriate, previous years.

Furthermore, supplementary sources are utilized to gather information for processes such as defining the target population, segmenting the sample, or measuring the similarity between experimental groups. These additional sources include:

- **Administrative records (Educa system):** managed by the Department of Education of Navarre, this source is used to collect basic data from the students and schools necessary for the design of the program and selection of eligible schools.
- **Questionnaire to families:** aimed to collect socio-demographic data from participating families such as educational level, occupation, and receipt of social benefits by parents.
- **Questionnaires to tutors:** gather demographic data post-intervention, covering previous experience as a tutor and teacher, social motivation, performance, satisfaction with the tutoring work, program usefulness, expected student outcomes, participation, potential behavioral issues among students, and coordination with educational staff.
- **Questionnaire to the participants schools:** aimed at management teams and tutors of the participating students, measures the satisfaction with the program.

3.4 Indicators

This section describes the indicators that this study used to evaluate the impact of the itinerary, divided by themes related to the hypotheses described above.

1. Mathematical results

To assess the main and secondary hypotheses of this section, four indicators are derived from the data obtained through mathematics tests and school report cards. These indicators include:

- **Standardized Math Test Score:** this indicator, designed to test the primary hypothesis, involves collecting students' scores from the math test. Scores range from 0 (all incorrect answers) to 10 (all correct answers).
- **School grades in mathematics:** constructed from students' report card data to test the main hypothesis, this indicator captures students' grades in mathematics. For the first wave, the average grade from the first and second quarters of the 2022-23 academic year is used as the pre-intervention reference, while the grade from the third quarter and final grades of the same academic year serves as the post-intervention observation. For the second wave, the grade from the third quarter of the 2022-23 academic year is taken as pre-intervention reference, and the grade from the first quarter of the subsequent academic year 2023-24 is taken as a post-intervention observation.[16] Grades from quarterly grades are on a scale of 0 to 10, while final grades are on a scale of 0 to 5. [17] To assess the intervention's effect on the final grade, it is compared with the average value of the first and second quarters of the 2022-23 academic year, normalized to a scale between 0 and 5.
- **School grades in mathematics for the following school year to the intervention**: this indicator, addressing the secondary hypothesis of long-term improvement in mathematics grades, is derived from the final grade in mathematics in the year following the intervention, scored out of 5. As a lagging indicator, the pre-intervention indicator coincides with that used in the previous indicator, namely the average of the first and second quarters of the 2022-23 academic year normalized to a scale between 0 and 5 for the first wave, and the final grade in mathematics of the 2022-23 academic year for the second wave. As of the date of this report, this indicator has not been evaluated.
- **Overall school grades**: constructed to test the second secondary hypothesis, this indicator represents the overall average grade of the student. Like the mathematics grade indicator, for the first wave, the overall grade from the first and second quarters of the 2022-23 academic year serves as the pre-intervention value, while the grade from the third quarter and final grades of the same academic year serves as the subsequent observation. For the second wave, the grade from the third quarter of the 2022-23 academic year is taken as the pre-intervention value, and the grade from the first quarter of the 2023-24 academic year is taken as a subsequent observation. Quarterly grades range from 0 to 10, while the final grades range from 0 to 5. The average of the first and second quarters of the 2022-23 academic year is captured both in its original scale and in its normalized value to a scale between 0 and 5 to facilitate comparison with the final grade of the school year.

[16] In the second wave, only the grade from the quarterly grades is considered due to the impossibility to capture the final year-end grades within the project pilot's assessment period.

[17] This scale corresponds to the following categories: "Not Passed" (0), "Insufficient" (1), "Sufficient" (2), "Good" (3), "Excellent" (4), and "Outstanding" (5).

2. Confidence in your math competencies

To gauge students' confidence in their mathematical competencies, indicators are derived from responses collected through a questionnaire. All indicators within this framework are standardized to have a mean of zero and a unit variance (standardized 0-1). The composition of these indicators is outlined below:

Self-Efficacy and Anxiety in Mathematics: this indicator is constructed from responses to 30 questions concerning students' relationship with mathematics, perceptions of classes, homework, exams, etc. Each of the 30 responses is rated on a scale of 1 to 5, where higher scores indicate greater self-efficacy and lower anxiety in mathematics. Following Anderson's methodology (2008), which aggregates information from variables attempting to measure a common latent variable, the indicator's value is derived. Essentially, this method calculates a weighted average of all variables, with weights determined by their correlations (lower correlations receive greater weight). The final indicator value is standardized with a mean of 0 and a standard deviation of 1. Additionally, this indicator is subdivided into four subscales: Self-Efficacy, Anxiety, Effort, and Interest in the teacher, each constructed and standardized according to Anderson's method.

Appreciation of the Subject of Mathematics: this indicator assesses students' affinity for the subject of mathematics based on their responses to a questionnaire item regarding their level of liking for mathematics. Responses to this question range from 1 (I like it a lot) to 5 (I don't like it). The indicator is derived from standardizing this variable to a range of 0-1.

3. Academic performance

To assess the hypothesis regarding academic performance, the following indicator has been utilized:

Time Spent Doing Homework: this indicator is derived from students' responses to a questionnaire item regarding the time they dedicate to homework. Responses are categorized on a scale of 1 to 7, representing different time intervals (1=Less than 15 minutes, 2=15-30 minutes, 3=30-60 minutes, 4=1-1.5 hours, 5=1.5-2 hours, 6=2-2.5 hours, 7=more than 2.5 hours). The indicator is generated by standardizing the variable, resulting in a mean of 0 and a standard deviation of 1.

At the time of this report, the information on unjustified absences necessary to construct an indicator corresponding to the main hypothesis in this section was not available.

4. Academic Aspirations

Also based on the information collected in the questionnaires administered to students, the following indicators are defined to assess the hypotheses of higher expectations of studies (main hypothesis) and the desire to attend university (secondary hypothesis), respectively:

Aspirations to complete baccalaureate: this indicator is measured through a questionnaire item asking students about their aspirations. If a student indicates a desire to pursue baccalaureate studies, the indicator is assigned a value of 1; otherwise, it receives a value of 0.

Desire to go to university: like the previous indicator, this measure is derived from a questionnaire item regarding students' intentions to pursue university education. If a student expresses a desire to attend university, the indicator is assigned a value of 1; otherwise, it is assigned a value of 0.

5. Wellbeing at school

Finally, to evaluate the hypotheses of greater well-being in school, several indicators are constructed from the information obtained through the questionnaire administered to the students at two different points in time. All indicators in this block are standardized to have a mean of zero and a unit variance (standardized 0-1). The following indicators are utilized:

School stress: This indicator comprises information obtained from 11 questions in the student questionnaire concerning school-related stress. The indicator synthesizes the responses to the questions using Anderson's method and is standardized with a mean of 0 and a standard deviation of 1.

Furthermore, the school stress indicator is segmented into three subscales that delineate the specific areas of impact of actions. To achieve this, small groups of responses are extracted from the 11 answers that constitute the overall indicator, resulting in independent subscales of Effort, Reward, and Overload. All the procedures maintain consistency with Anderson's methodology and standardization.

Well-being and motivation: Based on 14 questions from the student questionnaire addressing well-being and motivation, a composite indicator is constructed using the Anderson method. The scores are standardized to a range of 0-1. This indicator is only constructed for the initial wave[18].

Like the school stress indicator, the well-being and motivation indicator is divided into three subscales, enabling the specific impact areas of actions to be identified. For this purpose, small groups of responses are derived from the 14 responses comprising the global indicator, resulting in two independent subscales of Well-being and Motivation. These processes adhere to Anderson's methodology and are standardized with a mean of 0 and a standard deviation of 1.

GRIT socio-emotional skills: This indicator is constructed based on the interaction of 10 questions from the questionnaire that assess students' ability to confront and sustain their goals.

Locus of control: This indicator approximates students' attribution of effort perception and is constructed through interaction with 2 questionnaire questions.

Language subject appreciation: This is measured based on the information obtained from students' responses to a questionnaire question regarding their affinity for the Spanish or Basque language

[18] In the second wave, the questions pertaining to school well-being are removed from the questionnaire to reduce its length, thereby facilitating students' response to it.

subject. Responses to this question range from 1 to 5 (1=I like it a lot|2=I like it a lot|3=I don't like it or dislike it|4=I like it a little|5=I don't like it).

Life satisfaction: Constructed through the interaction of responses to 5 questions in the questionnaire regarding students' satisfaction with themselves, friends, family, school, and the environment. Like the well-being and motivation indicator, this indicator is solely constructed for the first wave.

3.5 Experiment design

To evaluate the impact of various interventions versus no intervention on indicators of mathematical outcomes, confidence in mathematical skills, academic performance, academic aspirations, and school well-being, an experimental assessment (RCT) is utilized, wherein participants are randomly assigned to either the control group or different treatment groups.

Recruitment of the beneficiaries of the intervention

The target population for the intervention comprises students in the third cycle of primary education and the first cycle of secondary education (from the 5th year of primary school to the 2nd year of ESO) attending public schools located in areas of Navarre with high vulnerability rates. The identification of this population is a collaborative effort between the Department of Education and the Department of Social Rights, as elaborated in section 2.1. The identification of the target population was completed with 49 schools in the first wave and 85 in the second.

The recruitment method involves contacting all identified schools, explaining the pilot project (including objectives, methodology, tasks, and deadlines), and inviting them to participate. Schools are responsible for initially accepting participation in the program and subsequently informing students and families about the opportunity to participate. Ultimately, it is the families of the students who decide and sign the informed consent.

After contacting the students and their families, thoroughly explaining the project[19], and obtaining informed consent, the sample of participant students is obtained. This process is followed in both waves of the pilot project.[20]

[19] Special emphasis is placed on the possibility of assigning participating schools and students in the program to a pure control group. This means that students continue to be part of the experiment but do not receive tutoring.

[20] In the first wave, as an incentive to participate, the possibility is offered that the students who are finally assigned to the control group can receive tutorials at the end of the experiment (during the second wave, but without being part of the sample of this one). In the recruitment process of the second wave, this incentive cannot be offered as there are no subsequent waves of tutoring.

Informed Consent

One of the fundamental ethical principles of research involving human beings (respect for persons) requires study participants to be informed about the research and consent to be included in the study. Informed consent is usually part of the initial interview and has two essential parts: the explanation of the experiment to the person, and the request and registration of their consent to participate. Consent should begin with a comprehensible presentation of key information that will help the person make an informed decision, i.e., understand the research, what is expected of it, and the potential risks and benefits. Documentation is required as a record that the process has taken place and as proof of informed consent, if so.

Informed consent is required in most research and may be oral or written, depending on different factors such as the literacy of the population or the risks posed by consent. Only under very specific circumstances, such as when the potential risks to participants are minimal and the informed consent is very complex to obtain or would harm the validity of the experiment, informed consent may be avoided, or partial information may be given to participants with the approval of the ethics committee.

Random assignment of participants

After obtaining consent and finalizing the sample selection for the study, participants in the experiment are randomly assigned to either the control group or various treatment groups. Randomization is pivotal in RCTs for establishing a causal relationship between treatment and outcomes. When conducted properly, this process ensures that both treatment and control groups are statistically comparable across observable and unobservable variables. This parity lays the groundwork for accurately measuring potential effects stemming from the intervention.

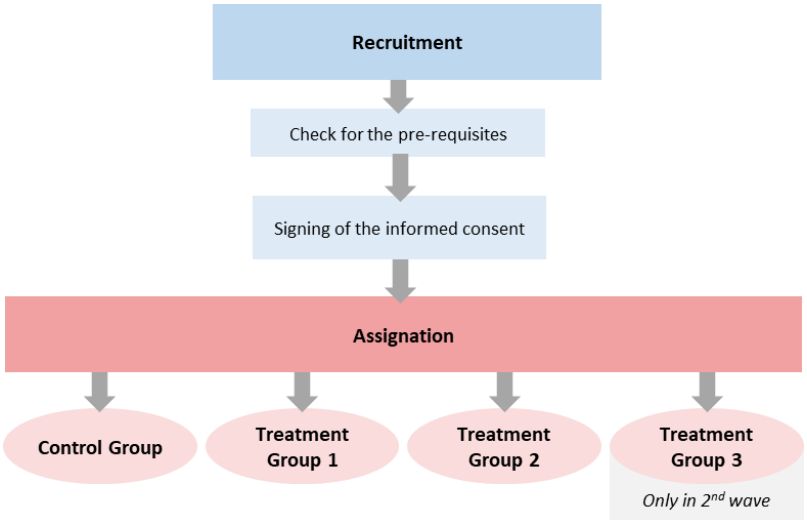
The experiment's design includes the random and equitable distribution of students participating in the pilot project across different control and treatment groups. Randomization occurs at the student level, employing a stratified approach to ensure balanced groups based on certain characteristics relevant to outcome indicators. This approach aims to prevent potential imbalances that could lead to erroneous measurements of intervention effects. Consequently, the sample is categorized into strata of participants sharing these characteristics, and random assignments are made within each stratum. Thus, from a design perspective, a balance is sought in the distribution of participants according to **school**, **school year** and **sex**, in the first wave, and **vehicular language** (Spanish or Basque) is added in the second wave. To this end, the sample is classified into strata of participants who share the characteristics and random assignment is made within each stratum.

It's important to note that the strata based on school, school year, and vehicular language are "strict," meaning each tutorial group must consist of students who share these characteristics. However, gender serves as a stratification variable to ensure gender balance in the experimental groups, allowing for the mixing of students of different sexes within tutoring groups.

Given the tutoring format (two to three students per tutor) and the stratification constraints, adjustments are necessary to allocate the appropriate number of students to each experimental group

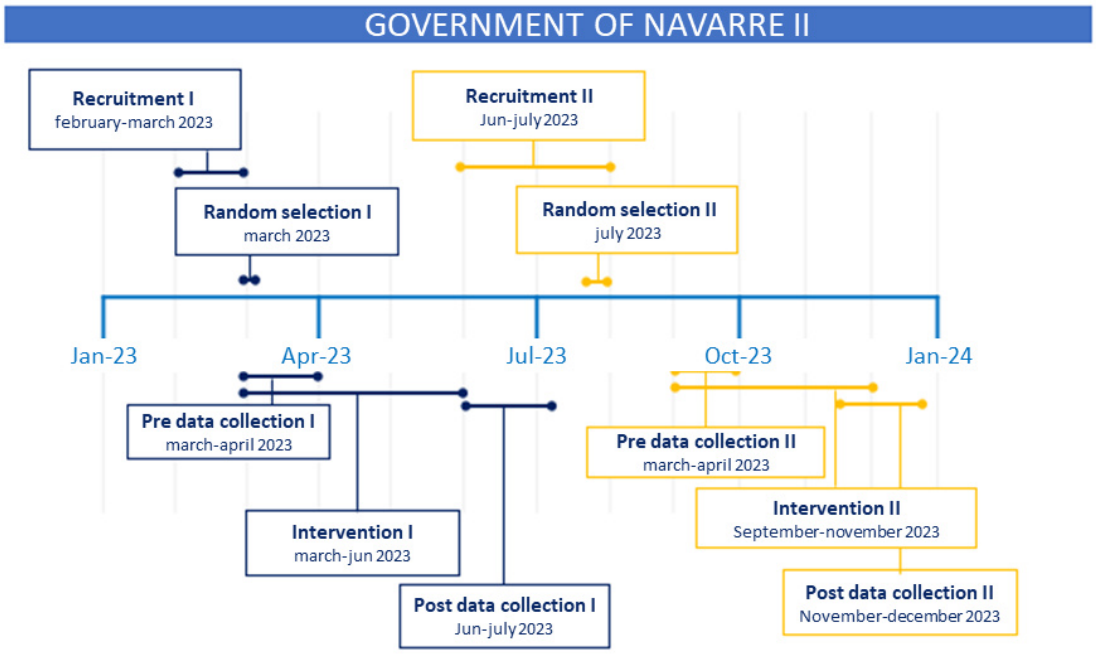
within each stratum. With this in mind, random assignment is conducted to the control and treatment groups, ensuring a suitable level of homogeneity in terms of basic characteristics across groups or strata.

Figure 4: Sample Design



The **Figure 5** It shows the time frame for the implementation and evaluation of the itinerary.

Figure 5: Evaluation timeline



4 Description of the implementation of the intervention

This section describes the practical aspects of how the intervention was implemented as part of the evaluation design. It details the results of the participant recruitment process and other relevant logistical aspects to contextualize the evaluation results.

4.1 Sample Description

As explained in previous sections and shown in **Table 1**, the initial approach to the sample involved selecting public schools that meet the defined vulnerability criteria. All eligible schools were invited to participate, and it was up to them to decide whether to join the program. This process led some schools to opt out, reducing the initial sample size. Although the target population is defined at the school level, the students themselves make up the experimental unit or sample of the pilot project. Participation is also subject to the decision of families or legal guardians, who give their consent by signing the informed consent form.

It is noteworthy that the sample size increased in the second wave compared to the first, growing from 49 eligible schools to 85, and from 545 students to 799 participants. This increase is due to lowering the at-risk-of-poverty rate required for schools to qualify as potential beneficiaries, as detailed in **Section 3.5** on the experiment design.

Table 1: Candidates and participant sample of the project

Concept	Sample from wave 1	Sample from wave 2
Public schools in vulnerable areas	49 schools	85 schools
∟ Public schools in vulnerable areas **accepting the invite to participate in the study.**	18 schools	39 schools
∟ Students whose families sign the **informed consent to participate in the program**	**545 students**	**799 students**

Characteristics of the final evaluation sample

From the sample of 1,344 students who provided informed consent (545 in wave 1 and 799 in wave 2), various factors resulted in incomplete data for some students. For instance, in the second wave,

751 out of 799 students took the mathematics level test, and only 451 completed the first survey[21]. Additionally, in some cases, information was missing because students did not answer specific questions in the questionnaire, or it was not possible to obtain data from their academic records.

Tables 2.1 and **2.2** present descriptive statistics for the sociodemographic variables and outcome indicators for the first and second waves, respectively. The tables are structured with six columns: variable name, sample size, mean, standard deviation, and minimum and maximum values.

The data in the tables come from students' administrative records, baseline surveys, level tests administered before the interventions, report cards, and the characteristics of the tutors assigned to the treatment groups. Therefore, the number of observations for each variable and outcome indicator varies, depending on the ability to capture information from different sources.[22]

For the first wave (as shown in **Table 2.1**), the analysis of the students' predetermined characteristics reveals considerable diversity among participants. Sixty-nine percent of students are enrolled in primary education (36% in fifth grade and 33% in sixth grade), and 31% are in secondary education (18% in the first year of ESO and 13% in the second year of ESO). Fifty percent of participants are female, the average age is 12.46 years, and 65% have Spanish nationality. Additionally, 66% are listed in the census of Specific Educational Support Needs. Geographically, 24% of the sample attends school in an urban area, 47% in a semi-urban area, and 29% in a rural area.

Regarding the Specific Educational Support Needs variable, 66% of participants are included in this census. Given the discretion of school counsellors in including students in this census, this statistic can be considered a minimum value.

Moving to the outcome indicators before the intervention, students in the first wave scored an average of 3.02 out of 10 on the mathematics level test, with the highest score being 8.4. On the other hand, the grade obtained from the participants' quarterly bulletins shows that the average school grade in mathematics in the first two quarters is 4.45 out of 10, and the overall average grade is 5.53.

Regarding the indicators constructed from the information from the baseline survey[23] in the assessment questions about language and mathematics subjects (where 1 means "I don't like it" and

[21] For the first wave, all students took the level test and responded to the survey. In the second wave, the initial survey was conducted after students were informed of their assignment to the treatment groups. Table A-1 in the appendix shows that participation rates are statistically different between the control group and the treatment groups. Table A-1 reports the results of linear regressions where the dependent variables are binary, equal to one if the student's data is missing for the initial survey in wave 2. The independent variables are binary, equal to one for assignment to each of the treatment groups, with the control group as the omitted category. It is observed that the higher participation in the initial survey compared to the control group is statistically significant at the 5% level for treatment group 1 and at the 1% level for treatment groups 2 and 3.

[22] For administrative data, the number of observations matches the number of participants. However, the information captured through surveys or exams is dependent on the response rate of students, family members, or guardians, resulting in a lower number of observations compared to the initial sample.

[23] Several of these indicators are not standardized in this description, for a better interpretation of them. Later, in the analysis of results, its standardized value is considered with mean 0 and standard deviation 1.

5 means "I like it a lot"), students reported an average score of 3.69 for language and 3.27 for mathematics. Students also reported spending an average of 2.71 on a 1 to 7 scale of time spent on homework. Regarding academic aspirations, 55% of students expressed a desire to continue studying in high school after finishing compulsory education, and 87% indicated they would like to attend university.[24]

The last set of indicators shows the average characteristics of the tutors responsible for the tutorials. In wave 1, the tutors were on average 34 years old, 82% were women, and 97% were born in Spain. Most tutors had previous teaching experience (91%), mostly in mathematics (84%), though only half had prior experience as tutors.

Table 2.1: Descriptive statistics of the sample in wave 1

Variable	N	Mean	Standard deviation	Minimum	Maximum
Predetermined Characteristics of Students					
5 Primary School	545	0.36	0.48	0	1
6 Primary School	545	0.33	0.47	0	1
1 Secondary School	545	0.18	0.38	0	1
2 Secondary School	545	0.13	0.34	0	1
Female	545	0.50	0.50	0	1
Age	545	12.46	1.20	11	16
Spanish Nationality	545	0.65	0.48	0	1
Specific Educational Support Needs Census	545	0.66	0.48	0	1
School in Urban Area	545	0.24	0.43	0	1
School in Semi-Dense Area	545	0.47	0.50	0	1
School in Rural Area	545	0.29	0.45	0	1
Outcome Indicators (Pre-Intervention)					
Score on Mathematics Level Test	520	3.02	1.62	0.45	8.40
School Grades in Mathematics (1st and 2nd Quarters) (0-10)	539	4.45	2.16	0	10
School Grades in Mathematics (1st and 2nd Quarters) (0-5)	539	1.80	1.21	0	5
Average School Grade (1st and 2nd quarters) (0-10)	540	5.53	1.48	0	9

[24] The indices of self-efficacy and anxiety in mathematics, school stress, life satisfaction, well-being, and motivation to go to school, socio-emotional skills, and locus of control are not commented on since they are indices composed of several questions of the questionnaire and standardized using the method proposed by Anderson (2008). This method aggregates information from a set of variables that attempt to measure a common latent variable. Intuitively, the method calculates a weighted average of all the variables, where the weight assigned to each of them depends on how correlated it is with the others (the lower the correlation, the greater the weight). Because it has no natural measures, the standardized indicator has been used to have a null mean and unit variance, which allows a better interpretation of the final results, but makes it difficult to interpret its descriptive statistics.

Variable	N	Mean	Standard deviation	Minimum	Maximum
Average School Grade (1st and 2nd Quarters) (0-5)	540	2.24	1.06	0	5
Self-Efficacy and Math Anxiety Index	405	0	1	-2.85	3.20
School Stress Index	485	0	1	-3.09	3.56
Life Satisfaction Index	506	0	1	-4.54	1.03
Well-Being and School Motivation Index	464	0	1	-3.69	1.81
Socio-Emotional Skills Index	461	0	1	-2.64	2.73
Locus of Control Index	498	0	1	-1.98	1.02
Appreciation Level of Spanish/Basque Language	514	3.69	1.11	1	5
Appreciation Level of Mathematics	519	3.27	1.25	1	5
Time Spent on Homework	510	2.71	1.25	1	7
Plan to Study in High School	519	0.55	0.50	0	1
Desire to Attend University	477	0.87	0.33	0	1
Pre-Determined Characteristics of Tutors					
Female	365	0.82	0.39	0	1
Age	365	33.92	7.38	24	51
Born in Spain	365	0.97	0.18	0	1
Master's Degree	365	0.66	0.47	0	1
Has Worked as a Teacher Previously	365	0.91	0.29	0	1
Has Worked as a Mathematics Teacher Previously	365	0.84	0.36	0	1
Has Worked as a Tutor Previously	365	0.50	0.50	0	1

Moving on to the second wave, **Table 2.2** reveals that 80% of students are enrolled in primary education (49% in fifth grade and 31% in sixth grade), while 20% are in secondary education (11% in the first year of ESO and 9% in the second year of ESO). Among them, 52% are female, with an average participant age of 10.77 years, and 69% hold Spanish nationality. Additionally, 59% are included in the census of Specific Educational Support Needs. Regarding school location, 11% of the sample attends an urban school, 24% attends a semi-urban school, and the remaining 65% attends a rural school.

In terms of academic performance, students in the second wave also obtained low scores on the initial level test in mathematics, averaging 3.01 out of 10, with the highest score being 9.2. Quarterly bulletins indicate that the average grade in mathematics for the third quarter of the previous year (2022-2023) is 2.65 out of 5, and the overall average grade for all subjects is 3.27 out of 5.

On the language and mathematics assessment questions, students in the second wave reported average scores of 3.80 and 3.63 out of 5, respectively, while the average result for the index of time spent on homework is 2.50 out of 7. As for academic aspirations, 75% of students plan to pursue high school, and 83% express a desire to attend university.

Regarding the average characteristics of tutors in wave 2, 69% and 70% of female tutors reported previous mentoring experience.

Table 2.2: Descriptive statistics of the sample in wave 2

Variable	N	Mean	Standard deviation	Minimum	Maximum
Predetermined Characteristics of Students					
5 Primary School	799	0.49	0.50	0	1
6 Primary School	799	0.31	0.46	0	1
1 Secondary School	799	0.11	0.32	0	1
2 Secondary School	799	0.09	0.28	0	1
Female	799	0.52	0.50	0	1
Age	799	10.77	1.25	9	15
Spanish Nationality	799	0.69	0.46	0	1
vehicular language (Spanish or Basque)	799	0.97	0.17	0	1
Specific Educational Support Needs Census	799	0.59	0.49	0	1
School in Urban Area	799	0.11	0.31	0	1
School in Semi-Dense Area	799	0.24	0.42	0	1
School in Rural Area	799	0.66	0.47	0	1
Outcome Indicators (Pre-Intervention)					
Score on Mathematics Level Test	751	3.01	1.70	0	9.20
School Grades in Mathematics (3rd Quarter 2022/23) (0-5)	796	2.65	1.48	0	5
Average School Grade for the Year (3rd Quarter 2022/23) (0-5)	797	3.27	0.96	0.78	5.00
Self-Efficacy and Math Anxiety Index	361	0	1	-2.79	3.31
School Stress Index	426	0	1	-3.24	3.00
Life Satisfaction Index	429	0	1	-4.70	1.00
Socio-Emotional Skills Index	428	0	1	-3.12	2.45
Locus of Control Index	449	0	1	-1.40	1.39
Appreciation Level of Spanish/Basque Language	448	3.80	1.04	1	5
Appreciation Level of Mathematics	450	3.63	1.25	1	5
Time Spent on Homework	12	2.50	1.62	1	7
Plan to Study in High School	12	0.75	0.45	0	1
Desire to Attend University	12	0.83	0.39	0	1
School Stress Index	426	3.22	1.18	-0.60	6.77
Pre-Determined Characteristics of Tutors					
Female	470	0.69	0.46	0	1
Age	470	32.46	6.28	24	51

Variable	N	Mean	Standard deviation	Minimum	Maximum
Born in Spain	470	0.97	0.17	0	1
Master's Degree	470	0.58	0.49	0	1
Has Worked as a Teacher Previously	470	0.94	0.23	0	1
Has Worked as a Mathematics Teacher Previously	470	0.86	0.35	0	1
Has Worked as a Tutor Previously	470	0.70	0.46	0	1

Based on the information reported from both waves, it is evident that they exhibit relatively similar results. However, some differences stand out. In the second wave, there was a notable increase of 13 percentage points (p.p.) in the presence of students in 5th grade accompanied by a decrease in the presence of students in 6th grade and 1st and 2nd year of ESO by 2, 7, and 4 p.p. respectively. Additionally, the presence of schools in rural areas increased by 36 p.p. in the second wave, compared to a decrease of 23 p.p. in semi-dense areas and 13 p.p. in urban areas.

Although marks in the mathematics test are very similar between the waves, curricular grades are approximately 1 point out of 5 higher for second-wave students. Furthermore, these students exhibit a slightly higher appreciation for language arts and mathematics, along with a greater desire to pursue high school. However, the time spent doing homework and the desire to attend college remain very similar between waves.

Regarding the profile of tutors, it remains very similar between waves, except for the percentage of tutors with previous experience giving tutorials, which increased from 50% to 70%. This increase is partly attributed to the fact that 16 tutors from wave 1 also participated in the second wave.

4.2 Random Assignment Results

To ensure the validity of the random assignment method described in **Section 3.5**, a balance test is conducted. This test aims to demonstrate that, on average, the observable characteristics of participants in different groups are equivalent. Achieving balance among experimental groups is crucial for accurately inferring the causal effect of the program through result comparisons.

In **Table 3**, the outcomes of randomization are illustrated. In the first wave (with a total of 545 students), randomization yielded 30% (164) of students in the control group, while 36% (195) and 34% (186) were allocated to treatment groups 1 and 2, respectively. In the second wave, out of 799 students, 26% (209) were assigned to the control group, 25% (198) to treatment group 1, 26% (204) to treatment group 2, and 24% (188) to treatment group 3.

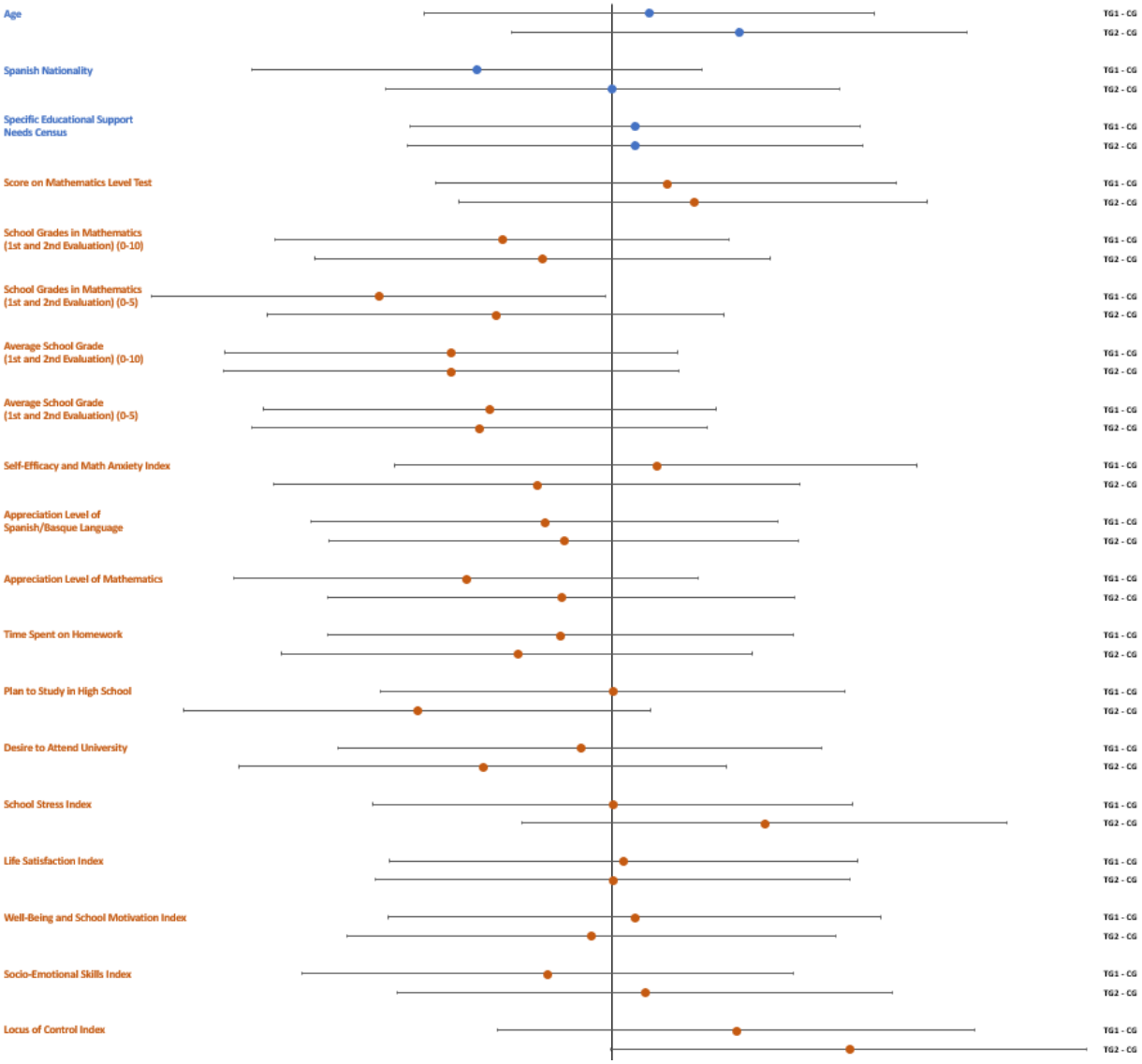
Table 3: Result of random assignment

Wave	Control Group	Treatment group 1	Treatment group 2	Treatment group 3	TOTAL
1	164	195	186	-	**545**
	30%	36%	34%	-	**100%**
2	209	198	204	188	**799**
	26%	25%	26%	24%	**100%**

To evaluate the comparability of treatment and control groups, balance tests are conducted on the variables collected during the initial survey. **Figures 6.1** and **6.2** depict the outcomes of these tests. Each observable variable's mean difference between treatment groups 1, 2, or 3 and the control group is represented by a data point, along with a 95% confidence interval. If the confidence interval encompasses zero (on the vertical axis), it indicates that the mean difference between groups lacks statistical significance, suggesting balance in that characteristic. Conversely, if the confidence interval does not include zero, it signifies a statistically significant difference, indicating imbalance.

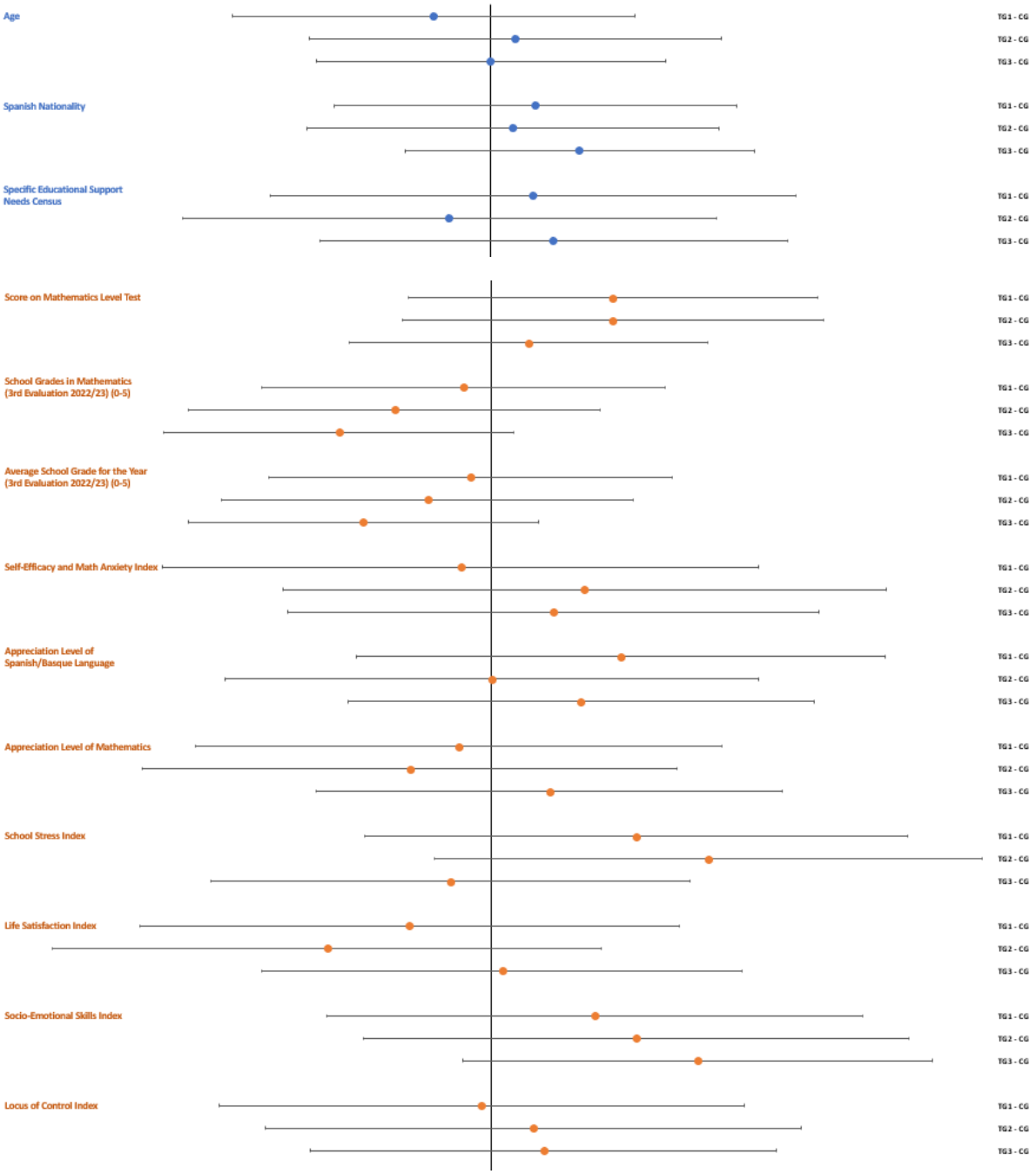
Both figures demonstrate that treatment and control groups exhibit no statistically significant differences across many variables. However, in the first wave an imbalance is observed in school grades for mathematics. Due to this statistically significant disparity, the regressions presented in the results section consistently adjust for the dependent variable's baseline survey value. This adjustment accounts for the possibility that control and treatment groups may not commence at equivalent levels.

Figure 6.1: Standardized mean difference between treatment and control groups (95% confidence interval) in wave 1



Note: the socio-demographic variables are shown in blue, and the indicators used for the evaluation of the project are shown in orange.

Figure 6.2: Standardized mean difference between treatment and control groups (95% confidence interval) in wave 2



Note: the socio-demographic variables are shown in blue, and the indicators used for the evaluation of the project are shown in orange.

4.3 Degree of participation and attrition by groups

The group that signed the informed consent represents the experimental sample that was randomly assigned to the control and treatment groups. However, participation in the program as well as response to the different surveys, baseline and endline questionnaires, and level tests are voluntary. For this reason, it is convenient to analyze the degree of participation in the program since the estimation of results will be sensible to the participation achieved. For example, if participation in a treatment activity is low, the students of that treatment and the control students will be very similar, and it will be more difficult to find an effect. On the other hand, this section tests whether the non-completion of the post-intervention surveys or level tests by some of the students reduces the comparability of the control and treatment groups between the different time points, if the response rate is different between groups or according to the demographic characteristics of the participants in each group.

Degree of participation

The process of randomizing participants reached a final sample of 545 students in the first wave and 799 students in the second wave, divided among the different experimental groups. The following table captures the attendance of students to the tutorials, both in the first wave and in the second, presenting the number of observations recorded, the mean, the standard deviation, the minimum, and the maximum of the sample. Thus, 76% of those assigned to the first wave missed at least one tutorial without justified cause. On average, they missed 4.66 sessions, and only 2% missed at least one session for good cause. Additionally, 6% in the first wave and 5% in the second wave were late for a mentoring session. On average, they were late for 0.14 tutorials, with 9 being the maximum number of tutorials a student arrived late to.

Regarding the second wave, 66% of students missed at least one tutorial without justified cause, and on average missed 3.79 sessions. Only one student missed a session for a good cause. 5% were late for a tutorial on at least one occasion, and, on average, they were late for 0.08 tutorials, with 5 being the maximum number of tutorials they were late for. Thus, attendance results are better in the second wave, highlighting the decrease in the percentage of students who missed tutorials, and the average number of tutorials missed.

Table 4: Absence from tutorials

	N	Mean	Standard deviation	Minimum	Maximum
First wave					
Absence	381	0.76	0.43	0	1
Number of Absences	381	4.66	4.98	0	19
Justified Absence	381	0.02	0.12	0	1
Number of Justified Absences	381	0.02	0.12	0	1
Late Arrival	381	0.06	0.24	0	1
Number of Times Arriving Late	381	0.14	0.75	0	9

Second Wave					
Absence	590	0.66	0.47	0	1
Number of Absences	590	3.79	4.98	0	18
Justified Absence	590	0.00	0.04	0	1
Number of Justified Absences	590	0.01	0.25	0	6
Late Arrival	590	0.05	0.22	0	1
Number of Times Arriving Late	590	0.08	0.42	0	5

Based on this, **Tables 5.1** and **5.2** provide an overview of the degree of participation in the intervention and sample attrition by experimental group. The content of these tables has been structured to show the total number of participants, by treatment group, who took the academic level test (1), for whom school record cards are available (2), who responded to the final survey (3), and for whom the survey conducted by their assigned tutor is available (only relevant to the treatment group).

In the first wave, it was observed that 10 students in the control group, 15 in treatment group 1, and 9 in treatment group 2 did not take the final test of mathematics. Additionally, the school report card is available for all students except in two cases, one in the control group and one in treatment group 1.[25] In the second wave, response rates are similar with 86% of the students of the control group completing the final test. The percentage of participation is higher among students assigned to treatment group 1 (95%), treatment group 2 (97%), and treatment group 3 (94%). The report card is available for all students except for three students in the control group and two students in each treatment group (in total, it is not available for 9 students).

It is also shown that 53% of the students in the control group in the first wave responded to the final survey, while in the treatment groups 69% (treatment group 1) and 66% (treatment group 2) did so. In the second wave, the share of completed final surveys is lower, with only 40% in the control group and 59%, 57%, and 60% in treatment groups 1, 2, and 3, respectively. For some indicators constructed with survey responses, the sample size is smaller because students did not necessarily answer all the points included in the survey.

Finally, **Tables 5.1** and **5.2** report in the last column the percentage of survey responses from tutors. This number should be zero for control groups, as it is pure control and has no tutor. For treatment groups, 7 students from treatment group 1 and 10 students from treatment group 2 could not be matched with the assigned tutor survey in the first wave. In the second wave, the rate of non-matching with the tutor survey is much higher. Of the 590 assigned to treatment, 20% could not be matched to the tutor's survey. Specifically, 48 students from treatment group 1, 55 students from treatment group 2, and 17 students from treatment group 3 could not be matched.

[25] In the specific case of the math grade, the report card is missing for three students.

Table 5.1: Experimental group dropout rate in wave 1

		Final Test (1)		School report card (2)		Final survey (3)		Tutor survey (4)	
		Obs.	%	Obs.	%	Obs.	%	Obs.	%
Control	Total	164	100%	164	100%	164	100%	164	100%
	Available	154	93.90%	163	99.39%	87	53.05%	0	0%
	Dropout	10	6.10%	1	0.61%	77	46.95%	164	100%
Treatment 1	Total	195	100%	195	100%	195	100%	195	100%
	Available	180	92.31%	194	99.49%	135	69.23%	188	96.41%
	Dropout	15	7.69%	1	0.51%	60	30.77%	7	3.59%
Treatment 2	Total	186	100%	186	100%	186	100%	186	100%
	Available	177	95.16%	186	100%	122	65.59%	176	94.62%
	Dropout	9	4.84%	0	0%	64	34.41%	10	5.38%
Total	Total	545	100%	545	100%	545	100%	545	100%
	Available	511	93.76%	545	100%	344	63.12%	364	66.79%
	Dropout	34	6.24%	0	0%	201	36.88%	181	33.21%

Table 5.2: Experimental group dropout rate in wave 2

		Final Test (1)		School report card (2)		Final survey (3)		Tutor survey (4)	
		Obs.	%	Obs.	%	Obs.	%	Obs.	%
Control	Total	209	100%	209	100%	209	100%	209	100%
	Available	180	86.12%	206	98.56%	83	39.71%	0	0%
	Dropout	29	13.88%	3	1.44%	126	60.29%	209	100%
Treatment 1	Total	198	100%	198	100%	198	100%	198	100%
	Available	189	95.45%	196	98.99%	116	58.59%	150	75.76%
	Dropout	9	4.55%	2	1.01%	82	41.41%	48	24.24%
Treatment 2	Total	204	100%	204	100%	204	100%	204	100%
	Available	198	97.06%	202	99.02%	116	56.86%	149	73.04%
	Dropout	6	2.94%	2	0.98%	88	43.14%	55	26.96%
Treatment 3	Total	188	100%	188	100%	188	100%	188	100%
	Available	178	94.68%	186	98.94%	112	59.57%	171	90.96%
	Dropout	10	5.32%	2	1.06%	76	40.43%	17	9.04%
Total	Total	799	100%	799	100%	799	100%	799	100%
	Available	745	93.24%	790	98.87%	427	53.44%	470	58.82%
	Dropout	54	6.76%	9	1.13%	372	46.56%	329	41.18%

Attrition by groups

To assess whether attrition rates vary significantly between the control group and the treatment groups, **Tables 6.1** and **6.2** present the results of linear regressions. In these regressions, the dependent variables are binary variables set to one if the student's data is missing for the primary data sources used in the study: the final math test (Tables 6.1 and 6.2, column 1), the school report card (only for the second wave, as there is minimal attrition in the first wave, Table 6.2, column 2), and the final survey (Tables 6.1, column 2, and Table 6.2, column 3).[26] The independent variables are binary variables set to one for assignment to each of the treatment groups, with the control group as the omitted category. In the first wave, it is noted that the lower attrition in the final survey in the treatment groups, compared to the control group, is statistically significant at 1%. In the second wave, the difference in attrition rates between the treatment groups and the control group is also statistically significant at 1% for both the final survey and the math test.

Table 6.1: Attrition rates by experimental groups of wave 1

	Did not take the final mathematics test (1)	Did not complete the final survey (2)
Treatment group 1	0.002	-0.167***
	(0.025)	(0.044)
Treatment group 2	-0.016	-0.143***
	(0.026)	(0.045)
Observations	545	545
R^2	0.221	0.385
Average of Control	0.061	0.470
Test T1=T2 p-value	0.473	0.581

Note: Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

[26] For now, no variables from the tutors' survey are incorporated, until the matching process is refined given the high number of students not matched to their tutors in the second wave.

Table 6Table .2: Attrition rates by experimental groups of wave 2

	Did not take the final mathematics test (1)	Unavailability of the school report card (2)	Did not complete the final survey (3)
Treatment group 1	-0.091***	-0.015	-0.188***
	(0.026)	(0.011)	(0.038)
Treatment group 2	-0.102***	-0.004	-0.181***
	(0.027)	(0.012)	(0.039)
Treatment group 3	-0.088***	-0.008	-0.190***
	(0.028)	(0.013)	(0.038)
Observations	799	799	799
R^2	0.352	0.259	0.627
Average of Control	0.139	0.014	0.603
Test T1=T2 p-value	0.621	0.182	0.856
Test T1=T3 p-value	0.890	0.512	0.954
Test T1=T2=T3 p-value	0.817	0.403	0.970

Note: Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Going deeper into the issue of group attrition, **Tables 7.1** and **7.2** expand upon the linear regressions from the previous two tables by introducing interactions between the treatment indicators and predetermined student characteristics (secondary education, sex, and urban area of residence). This allows for an analysis of whether the attrition rate correlates with any of these variables. In the first wave, no significant difference in attrition is observed, at least at the 10% significance level. In the second wave, there is a greater difference in attrition rates in the final survey between treatment group 2 and the control group for secondary school students, and in the probability of taking the math tests between treatment group 3 and the control group for female students. Both differences are statistically significant at 5%.

Table 7.1: Attrition selectivity among treatment groups in wave 1

	Did not complete the final survey	
Control variable	**Treatment group 1**	**Treatment group 2**
Secondary school	-0.043	-0.040
	(0.096)	(0.097)
Female	-0.061	0.062
	(0.089)	(0.090)
Urban	-0.029	-0.026
	(0.096)	(0.099)
Observations	545	545
Average of control	0.470	0.470

Note: To simplify the table, only the coefficients associated with the interactions between treatment and each control variable are presented in the same column for all the regressions performed, identified by a separation line. Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 7.2: Attrition selectivity among treatment groups in wave 2

	Did not take the final mathematics test			Did not complete the final survey		
Variable control	**Treatment group 1**	**Treatment group 2**	**Treatment group 3**	**Treatment group 1**	**Treatment group 2**	**Treatment group 3**
Secondary school	-0.017	-0.074	-0.052	0.124	-0.189**	-0.063
	(0.065)	(0.052)	(0.079)	(0.088)	(0.075)	(0.110)
Female	0.029	-0.026	-0.178**	0.107	-0.125	-0.064
	(0.071)	(0.053)	(0.088)	(0.091)	(0.077)	(0.102)
Urban	0.076	-0.045	-0.136	0.012	-0.121	-0.022
	(0.082)	(0.056)	(0.088)	(0.096)	(0.077)	(0.116)
Observations	799	799	799	799	799	799
Average of control	0.139	0.139	0.139	0.603	0.603	0.603

Note: To simplify the table, only the coefficients associated with the interactions between treatment and each control variable are presented in the same column for all the regressions performed, identified by a separation line. Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

To assess whether the level of treatment adherence differs significantly between treatment groups, **Tables 8.1** and **8.2** present the results of linear regressions. In these regressions, the dependent variables are as follows: a binary variable set to one if the student missed a tutorial without justified cause (column 1); the number of tutorials the student missed without just cause (column 2); a binary variable set to one if the student missed a tutorial for good cause (column 3); the number of tutorials the student missed for good cause (column 4); a binary variable set to one if the student was late for a tutorial (column 5); and the number of tutorials for which the student was late (column 6). The independent variables are binary variables set to one for assignment to each of the treatment groups, with treatment group 1 as the omitted category.

In the first wave, it was observed that there were no statistically significant differences between the treatment groups in adherence to the tutorials. However, it is noted that students who participate in treatment group 2, i.e., in tutorials with three students, are on average more likely to be late for at least one tutorial, and on average to be late for more tutorials. These differences are statistically significant at 1% and 5% levels respectively.

In the second wave, students assigned to the treatment group with two students and a social-emotional component missed 1.17 more tutorials than students assigned to the treatment group with two students. This difference is statistically significant at 5%.

Table 8.1: Lack of attendance in wave 1

	Lack of attendance					
	Unjustified Absences		Justified Absences		Lateness	
	At any time (1)	Total (2)	At any time (3)	Total (4)	At any time (5)	Total (6)
Treatment group 2	0.045	-0.226	0.011	0.011	0.068***	0.204**
	(0.043)	(0.485)	(0.007)	(0.007)	(0.025)	(0.085)
Observations	381	381	381	381	381	381
Average of T1	0.756	4.661	0.016	0.016	0.060	0.139

Note: Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 8.2: Lack of attendance in wave 2

	Lack of attendance					
	Unjustified Absences		Justified Absences		Lateness	
	At any time (1)	Total (2)	At any time (3)	Total (4)	At any time (5)	Total (6)
Treatment group 2	0.034	0.399	0.005	0.032	0.006	0.054
	(0.047)	(0.501)	(0.006)	(0.035)	(0.025)	(0.046)
Treatment group 3	0.074	1.169**	-0.000	-0.000	-0.011	0.012
	(0.050)	(0.484)	(0.002)	(0.010)	(0.020)	(0.033)
Observations	590	590	590	590	590	590
Average of T1	0.658	3.795	0.002	0.010	0.051	0.083

Note: Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

5 Results of the evaluation

Random assignment of the experimental sample to the control and treatment groups ensures that, with a sufficiently large sample, the groups are statistically comparable. Therefore, any differences observed after the intervention can be causally associated with the treatment. Econometric analysis provides this comparison while allowing for the inclusion of other variables to improve the accuracy of the estimates and provides confidence intervals for these estimates. This section presents the econometric analysis and the estimated regressions, as well as the analysis of the results obtained.

5.1 Description of the econometric analysis: estimated regressions

The regression model specified to estimate the causal effect in a randomized experiment typically involves simply comparing the variable of interest between the treatment group and the control group, as randomization ensures statistical comparability. Given some previously documented imbalances and significant differences in survey participation rates, this analysis has been carried out

always controlling for the lagged value of the dependent variable (i.e., its value before the intervention). This control ensures that pre-intervention differences between the treatment and control groups are accounted for in the analysis.

The Intention-To-Treat (ITT) effect of the tutorials on the outcome variables is estimated. Specifically, the regression to be estimated in the first wave is:

$$y_i = \alpha + \beta_1 T1_i + \beta_2 T2_i + \gamma X_i + \delta y_{-1,i} + \epsilon_i$$

Where (y_i) is the dependent variable observed after the intervention for student (i); ($T1_i$) and ($T2_i$) are binary variables equal to one if the student is assigned to treatment group 1 (tutorials with 2 students) or treatment group 2 (with 3 students), respectively, with the control group being the omitted category; ($y_{-1,i}$) is the lagged value of the dependent variable (before the intervention); (X_i) contains the stratification variables; and (ϵ_i) is the error term robust to heteroscedasticity. In cases where the lagged value ($y_{-1,i}$) is not available, a binary variable equal to one will be included to indicate this, and ($y_{-1,i}$) will be replaced with zero. Linear regressions are estimated for all specifications, regardless of whether the results are continuous or discrete.

In the second wave, the specification of the regressions is identical to that in the first wave, except that a third binary variable ($T3_i$) is added to capture assignment to treatment group 3 (tutorials with 2 students plus the socio-emotional component).

The stratification variables used in the first wave are school, sex, and grade. In the second wave, they are school, sex, grade, and language (Spanish or Basque).

5.2 Analysis of the results

5.2.1 Primary and secondary outcomes

This section presents the results of the assessment on the primary and secondary indicators. All the indicators constructed with many variables (composite indicators) are standardized to have a mean of zero and a standard deviation of one, allowing all regression coefficients to be interpreted in terms of standard deviations, which helps when comparing effect sizes across different domains.

For each dependent variable, the regression results are reported in two columns. The first column shows the results uncontrolled for the value of the dependent variable before the intervention, and the second adds the lagged value as a control to the regression.

1. Mathematical results

Tables 9.1 and **9.2** show the results of the interventions on students' academic outcomes for waves 1 and 2, respectively. Both tables present regressions with three dependent variables: the final math test (columns 1 and 2), school grade in math, and overall school grade.

In **Table 9.1**, the results of the third quarter for both the mathematics school grades and the overall grade are presented using the original scale of 0 to 10 (columns 3 and 4, and 7 and 8), and the result

of the final grades on the original scale of 0 to 5[27] (columns 5 and 6, and 9 and 10). In both cases, the lagged value is the mean result of the first and second quarters on the original scale of 0 to 10 or standardized to a scale of 0 to 5 to be comparable to the dependent variable. All grades in the first wave correspond to the 2022/2023 academic year.

In **Table 9.2**, the dependent variables for school grades are the results of the first quarter (academic year 2023/2024) on the available scale from 0 to 10. The lagged value corresponds to the result of the third quarter in the 2022/2023 academic year on the available scale from 0 to 10.

In the first wave, **Table 9.1** shows no statistically significant effect on the math test score among students assigned to treatment group 1, but a positive effect of between 0.28 and 0.32 points in grade improvement for students assigned to treatment group 2. This positive effect is statistically significant at 10%. The coefficient for treatment group 1 is minimally different from the coefficient for treatment group 2 (p-values of 0.10 in both definitions). Regarding school grades in mathematics, a statistically significant, positive effect is detected for students assigned to treatment group 1 in the third quarter (column 4). A significant effect is observed in the final mathematics grade for the 2022/2023 academic year of 0.15 (treatment group 1) and 0.13 (treatment group 2); these effects are not statistically different from each other. Finally, no effects are detected in the overall grade for the academic year, either in the third quarter or the final grade.

Table 9.1: Effects on academic grades in wave 1

			Math school grade				Overall school grade			
	Math test		3Q (0-10)		Final (0-5)		3Q (0-10)		Final (0-5)	
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Treatment group 1	0.06	0.05	-0.08	0.21*	-0.12	0.15**	-0.17	0.09	-0.17	0.05
	(0.16)	(0.14)	(0.23)	(0.13)	(0.12)	(0.06)	(0.15)	(0.05)	(0.15)	(0.09)
Treatment group 2	0.32*	0.28*	0.04	0.05	0.09	0.13**	-0.04	0.07	-0.04	0.02
	(0.17)	(0.15)	(0.23)	(0.13)	(0.12)	(0.06)	(0.15)	(0.06)	(0.15)	(0.09)
Observations	511	511	542	542	543	543	543	543	543	543
Mean control	3.39	3.39	4.63	4.63	2.04	2.04	5.81	5.81	5.81	5.81
Test T1=T2 p.value	0.10	0.10	0.59	0.18	0.08	0.73	0.39	0.77	0.39	0.79
Strata	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Lag	No	Sí	No	Sí	No	Sí	No	Sí	No	Sí

Note: Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

In the second wave, **Table 9.2** shows that treatment group 1 has a statistically significant effect on the result of the mathematics test, with an effect between 0.4 and 0.6 points statistically significant at 5% (column 1) and 1% (column 2). The mean effects of the other two treatments are not statistically significant, although the coefficients are positive. In the specification with controls for the lagged value, the effect of treatment group 2 is not statistically different from treatment group 1 (p-value

[27] This scale corresponds to the categories: not yet achieved (0), insufficient (1), sufficient (2), good (3), very good (4), and outstanding (5).

0.14). As in the first wave, statistically significant effects on school math scores are observed in all treatments. Specifically, in the first quarter, the main specification with controls for the lagged value shows an effect of 0.29 (Treatment group 1), 0.32 (Treatment group 2), and 0.22 (Treatment group 3) points. The results for the specification without controls for the lagged value are similar, except for treatment group 3, consistent with the imbalance shown in **Table 15.2**. Although not statistically significant, the increase in statistical power is evident when controlling for the lagged variable. In the second wave, the effects of tutorials transfer to the average grade of all subjects. This is mechanically expected as we only take the marks of the first quarter to calculate the average, having detected an effect in mathematics.[28]

Table 9.2: Effects on academic grades in wave 2

	Math test		Math school grade 1Q (0-10)		Overall school grade Final (0-5)	
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Treatment group 1	0.61***	0.40**	0.22	0.29***	0.07	0.09*
	(0.21)	(0.19)	(0.16)	(0.11)	(0.10)	(0.04)
Treatment group 2	0.28	0.16	0.14	0.32***	0.03	0.10**
	(0.21)	(0.19)	(0.15)	(0.11)	(0.09)	(0.05)
Treatment group 3	0.23	0.13	-0.01	0.22**	-0.07	0.04
	(0.21)	(0.18)	(0.16)	(0.11)	(0.10)	(0.05)
Observations	745	745	759	759	759	759
Mean control	3.01	3.01	2.73	2.73	3.13	3.13
Test T1=T2 p.value	0.09	0.14	0.58	0.74	0.63	0.80
Test T1=T3 p.value	0.04	0.08	0.14	0.51	0.17	0.29
Test T1=T2=T3 p.value	0.10	0.17	0.33	0.62	0.39	0.42
Strata	Sí	Sí	Sí	Sí	Sí	Sí
Lag	No	Sí	No	Sí	No	Sí

Note: Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

2. Confidence in math skills and well-being at school

Tables 10.1 and **10.2** show the effects of tutoring on the index of self-efficacy and anxiety in mathematics (columns 1 and 2), appreciation of the subject of mathematics (columns 3 and 4), school stress index (columns 5 and 6), and the scale measuring the time students report doing homework each day for the first and second waves (columns 7 and 8). All variables are standardized, allowing the coefficients to be interpreted in terms of standard deviations.

[28] The regressions in columns 5 and 6 of Table 14 have been limited to only including students for whom the mathematics grade value is available in the school report card. The number of observations may be increased in the future when the rest of the grades have been recorded in the educational system of Navarre and shared for research purposes.

Table 10.1 shows no effect on the index of self-efficacy and anxiety in mathematics among students assigned to treatment group 1 in the first wave, but a positive effect of between 0.34 and 0.41 standard deviations for students assigned to treatment group 2. This positive effect is statistically significant at 5% when not controlled for the lagged value of the dependent variable before the intervention, and at 1% when controlling for the lagged value. The coefficient for treatment group 1 is statistically different from the coefficient for treatment group 2 (p-values 0.05 and 0.01). No statistically significant effects of the treatments are observed on the mathematics appreciation variable, although the coefficients are positive. Column 6 shows that treatment group 2 decreases school stress by 0.27 standard deviations, a significant effect at 5%. This effect is not statistically significant when the lagged value of the dependent variable is not included. Again, no statistically significant effects are observed among students assigned to treatment group 1. No statistically significant effects of any treatments are found on the time students report spending on homework.

Table 10.1: Effects on math confidence, stress, and performance in wave 1

	Mathematics							
	Self-efficacy and anxiety		Appreciation		School stress		Time spent on homework	
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Treatment group 1	0.04	0.05	0.11	0.16	0.03	-0.00	0.14	0.21
	(0.15)	(0.14)	(0.13)	(0.10)	(0.15)	(0.13)	(0.14)	(0.14)
Treatment group 2	0.34**	0.41***	0.04	0.15	-0.18	-0.27**	0.08	0.15
	(0.143)	(0.134)	(0.139)	(0.113)	(0.143)	(0.129)	(0.143)	(0.140)
Observations	280	280	343	343	326	326	343	343
Mean control	-0.12	-0.12	-0.03	-0.03	0.03	0.03	-0.09	-0.09
Test T1=T2 p.value	0.05	0.01	0.58	0.87	0.13	0.03	0.66	0.64
Strata	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Lag	No	Sí	No	Sí	No	Sí	No	Sí

Note: Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 10.2 shows the results for the second wave. Columns 1 and 2 show a positive effect of all treatments on the self-efficacy and anxiety index, although only a statistically significant effect is found at 5% (controlling for the lagged value of the dependent variable) and at 1% (when not) among students assigned to treatment group 3, which includes the socio-emotional component. The bottom panel of **Table 10.2** shows the p-values for contrasts of equality between the coefficients of the treatment variables. The results show that the hypothesis that the coefficients are equal cannot be rejected at any level of significance.

Columns 3 and 4 show a positive and statistically significant effect at 10% among students assigned to treatment group 3. Treatment groups 1 and 2 have no significant effect on the math appreciation index. Finally, no statistically significant effects of any of the three second-wave treatment groups are found on the school stress index or on the time students report spending on homework.

Table 10.2: Effects on math confidence, stress, and performance in wave 2

	Mathematics							
	Self-efficacy and anxiety		Appreciation		School stress		Time spent on homework	
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Treatment group 1	0.30	0.27	0.13	0.15	-0.17	-0.20	-0.05	-0.07
	(0.19)	(0.18)	(0.16)	(0.14)	(0.15)	(0.15)	(0.16)	(0.16)
Treatment group 2	0.19	0.14	-0.03	0.00	0.08	0.00	-0.04	-0.05
	(0.19)	(0.17)	(0.17)	(0.15)	(0.16)	(0.17)	(0.16)	(0.17)
Treatment group 3	0.47**	0.48***	0.27*	0.255*	-0.19	-0.15	0.14	0.12
	(0.193)	(0.176)	(0.159)	(0.141)	(0.153)	(0.145)	(0.157)	(0.160)
Observations	336	336	426	426	397	397	424	424
Mean control	-0.22	-0.22	-0.09	-0.09	0.09	0.09	-0.10	-0.10
Test T1=T2 p.value	0.53	0.42	0.25	0.23	0.08	0.13	0.94	0.92
Test T1=T3 p.value	0.33	0.23	0.25	0.32	0.90	0.69	0.16	0.17
Test T1=T2=T3 p.value	0.28	0.14	0.08	0.11	0.13	0.30	0.31	0.32
Strata	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Lag	No	Sí	No	Sí	No	Sí	No	Sí

Note: Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Overall, the tutorials in both waves have improved the indicator of students' confidence in their mathematical ability, except for the students in the first wave assigned to tutorials with 2 students (treatment group 1). This has likely increased their appreciation for the subject. On the other hand, the socio-emotional component of treatment group 3 does not improve the student's well-being compared to the other treatments. Only in the first wave is a reduction in school stress observed among students assigned to treatment group 2, where positive results were also observed in the self-efficacy and anxiety indicator.

The self-efficacy and anxiety indicator is constructed based on 30 questions grouped into four subcategories: self-efficacy, anxiety, effort, and perception of teacher interest. Columns 1-4 of **Tables A-2** and **A-3** in the appendix show the results for these subcategories using the main specification (controlling for the lagged value of the dependent variable) for the first and second waves, respectively. The results show that, in the first wave, the improvements in the aggregate indicator of self-efficacy and anxiety for treatment group 2 are due to better responses to questions related to self-efficacy, effort in the subject of mathematics, and the student's perception of teacher support and interest. In these last two variables, a positive and statistically significant impact is also observed among the students assigned to treatment group 1. For the second wave, the results are very similar. The effect on the aggregate indicator of self-efficacy and anxiety is mainly explained by an improvement in the confidence component for students assigned to treatment groups 1 and 3, and an improved perception of teacher support. No statistically significant effects are observed among students assigned to treatment group 2.

Columns 5-7 of **Tables A-2** and **A-3** show the results for the three subcategories that make up the school stress index: effort, reward, and overload (or overcommitment). In this case, no single component has more weight than another. The appendix (Tables A-4 and A-5) shows that there are no effects on other outcome indicators related to socio-emotional skills, locus of control, social well-being, and appreciation of the language subject.

3. Academic Aspirations

Tables 11.1 and **11.2** show the effects of tutoring on students' academic aspirations, measured by whether the student plans to study high school after compulsory education (columns 1 and 2) and whether they would like to attend university (columns 3 and 4). In both waves and indicators, the interventions have not had on average any statistically significant effect at the 10% level.

Table 11.1: Effects on academic aspirations in wave 1

	Baccalaureate		University	
Variable	**(1)**	**(2)**	**(3)**	**(4)**
Treatment group 1	0.07 (0.07)	0.01 (0.07)	0.07 (0.07)	0.06 (0.07)
Treatment group 2	-0.01 (0.07)	-0.05 (0.07)	0.06 (0.07)	0.04 (0.07)
Observations	342	342	342	342
Mean control	0.58	0.58	0.67	0.67
Test T1=T2 p.value	0.20	0.34	0.92	0.75
Strata	Sí	Sí	Sí	Sí
Lag	No	Sí	No	Sí

Note: Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 11.2: Effects on academic aspirations in wave 2

	Baccalaureate		University	
Variable	**(1)**	**(2)**	**(3)**	**(4)**
Treatment group 1	0.09	0.09	0.07	0.08
	(0.076)	(0.078)	(0.064)	(0.064)
Treatment group 2	0.03	0.03	0.03	0.04
	(0.08)	(0.08)	(0.07)	(0.07)
Treatment group 3	-0.01	-0.01	-0.10	-0.08
	(0.08)	(0.08)	(0.07)	(0.07)
Observations	426	426	424	424
Mean control	0.60	0.60	0.75	0.75
Test T1=T2 p.value	0.42	0.42	0.43	0.43
Test T1=T3 p.value	0.22	0.22	0.01	0.01
Test T1=T2=T3 p.value	0.44	0.45	0.03	0.03
Strata	Sí	Sí	Sí	Sí
Lag	No	Sí	No	Sí

Note: Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

5.2.2 Heterogeneity analysis

This section presents the analysis of the heterogeneity of effects according to participant characteristics. Specifically, it examines whether the effects on mathematics results and self-efficacy in this subject differ by sex, nationality, place of residence (urban or rural and semi-urban), academic year (primary or secondary), and previous level of mathematics (binary variable equal to one if the student's school grade in mathematics was above the median of the sample, calculated separately for primary and secondary). To do this, regressions like those in the previous section are specified, but they add the variable for which the heterogeneous effects are to be estimated, as well as the interaction of this variable with the treatment variables.

Tables 12.1 and **13.1** show the heterogeneous results for the math level test and school grades in the first wave. The coefficients of interest are those corresponding to the interaction between the treatment variables and the binary variable that includes the characteristics of the participants being analyzed (sex, nationality, place of residence, academic year, and previous level of mathematics).

The results show no heterogeneous effects on mathematics performance for any of the analyzed characteristics among students assigned to the tutorials with 2 students in the first wave (treatment group 1). However, positive, and statistically significant effects were found at 5% among students assigned to treatment group 2 (tutorials with 3 students) for students living in urban centers; the differential effect is 0.68 points and significant at 5%. Heterogeneous effects are also found for students with previous above-average math scores assigned to treatment group 2. On average, the

level test score for this group of students is 0.5 points higher than both students with below-average grades in the same treatment group and compared to the control group. On the other hand, when analyzing the impact of tutorials on school grades, only the coefficient of interaction between being assigned to the tutorials of 3 students (treatment group 2) and having Spanish nationality is positive and significant at 10%.

Table 12.1: Heterogeneous effects on the mathematics level test score in wave 1

	Standardized math test score				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
TG1	-0.005	0.060	-0.065	0.193	-0.109
	(0.199)	(0.188)	(0.156)	(0.175)	(0.178)
TG1 x female	0.103				
	(0.288)				
TG1 x Spanish nationality		-0.019			
		(0.213)			
TG1 x urban			0.460		
			(0.364)		
TG1 x secondary school				-0.471	
				(0.302)	
TG1 x prev. grade >50					0.318
					(0.211)
TG2	0.218	0.370*	0.108	0.408**	-0.017
	(0.207)	(0.219)	(0.173)	(0.174)	(0.188)
TG2 x female	0.114				
	(0.298)				
TG2 x Spanish nationality		-0.136			
		(0.230)			
TG2 x urban			0.681**		
			(0.341)		
TG2 x secondary school				-0.420	
				(0.328)	
TG2 x prev. grade >50					0.500**
					(0.216)
Constant	2.153***	2.145***	2.146***	2.157***	2.245***
	(0.172)	(0.175)	(0.171)	(0.172)	(0.175)
Observations	511	511	511	511	511
Control mean	3.386	3.386	3.386	3.386	3.386
TG1 + TG1*X	0.10	0.04	0.40	-0.28	0.21
TG2 + TG2*X	0.33	0.23	0.79***	-0.01	0.48***

Note: Strata-controlled regressions and baseline indicator. Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 13.1: Heterogeneous effects on the school mathematics grade in wave 1

	School grades in mathematics (third quarter of 2022-23)				
	(1)	(2)	(3)	(4)	(5)
TG1	0.249	0.371**	0.273*	0.203	0.042
	(0.178)	(0.175)	(0.141)	(0.145)	(0.179)
TG1 x female	-0.070				
	(0.252)				
TG1 x Spanish nationality		-0.259			
		(0.179)			
TG1 x urban			-0.226		
			(0.307)		
TG1 x secondary school				0.047	
				(0.283)	
TG1 x prev. grade >50					0.317
					(0.196)
TG2	0.122	-0.179	0.177	-0.052	-0.073
	(0.187)	(0.194)	(0.145)	(0.145)	(0.196)
TG2 x female	-0.136				
	(0.260)				
TG2 x Spanish nationality		0.343*			
		(0.206)			
TG2 x urban			-0.498		
			(0.313)		
TG2 x secondary school				0.324	
				(0.298)	
TG2 x prev. grade >50					0.209
					(0.205)
Constant	0.695***	0.702***	0.684***	0.689***	0.835***
	(0.189)	(0.192)	(0.185)	(0.188)	(0.221)
Observations	542	542	542	542	542
Control mean	4.632	4.632	4.632	4.632	4.632
TG1 + TG1*X	0.18	0.11	0.05	0.25	0.36**
TG2 + TG2*X	-0.01	0.16	-0.32	0.27	0.14

Note: Strata-controlled regressions and baseline indicator. Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 14.1 shows the heterogeneous results for the index of self-efficacy and anxiety in mathematics for the first wave. Heterogeneous effects are found for several groups. A positive and statistically significant effect was observed for students in both treatments, tutorials with 2 and 3 students, for students in secondary school who had previous grades in mathematics above the average. The last two lines of the table show the total effect of each treatment on these groups. Among high school students assigned to treatment group 1 and those assigned to treatment group 2, the math self-efficacy and anxiety index increased by 0.45 and 0.89 standard deviations, respectively (significant effects at 10% and 1%). Being assigned to treatment group 2 (tutorials with 3 students) also increases

the self-efficacy index among primary school students, but the effect is smaller, 0.59 standard deviations lower; this difference is significant at 10%. Finally, students with an above-average level of mathematics prior to the intervention also see their confidence in their mathematics performance improve. The effects are 0.39 standard deviations (significant at 5%) among students assigned to treatment group 1 and 0.55 standard deviations (significant at 1%) among students assigned to treatment group 2. No positive effects were observed for students with below-average math scores, but it was observed that among these students, those who were assigned to the tutoring sessions of treatment group 1 saw a decrease in their self-efficacy and anxiety index.

Table 14.1: Heterogeneous effects on mathematics self-efficacy in wave 1

	Self-efficacy in mathematics				
	(1)	(2)	(3)	(4)	(5)
TG1	0.032	0.042	0.171	-0.056	-0.408**
	(0.225)	(0.181)	(0.153)	(0.161)	(0.176)
TG1 x female	0.002				
	(0.288)				
TG1 x Spanish nationality		0.003			
		(0.212)			
TG1 x urban			-0.402		
			(0.322)		
TG1 x secondary school				0.506*	
				(0.299)	
TG1 x prev. grade >50					0.794***
					(0.195)
TG2	0.610***	0.243	0.400***	0.298*	0.189
	(0.198)	(0.208)	(0.152)	(0.153)	(0.184)
TG2 x female	-0.364				
	(0.269)				
TG2 x Spanish nationality		0.253			
		(0.227)			
TG2 x urban			0.042		
			(0.313)		
TG2 x secondary school				0.594*	
				(0.308)	
TG2 x prev. grade >50					0.361*
					(0.201)
Constant	-0.241**	-0.249**	-0.250**	-0.266***	-0.253**
	(0.099)	(0.100)	(0.099)	(0.097)	(0.100)
Observations	280	280	280	280	280
Control mean	-0.123	-0.123	-0.123	-0.123	-0.123
TG1 + TG1*X	0.03	0.04	-0.23	0.45*	0.39**
TG2 + TG2*X	0.25	0.50***	0.44	0.89***	0.55***

Note: Strata-controlled regressions and baseline indicator. Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Tables 12.2 and **13.2** show the heterogeneous results for the second wave with respect to mathematics grades. The results indicate that there are heterogeneous effects on mathematics performance, but these are not consistent across indicators. Table 12.2 shows positive and statistically significant results in the mathematics level test among students assigned to treatment group 1 who are women, of Spanish nationality, reside in urban areas, or have an above-average level of mathematics prior to the intervention. Among students assigned to treatment group 2, heterogeneous effects are only observed in those students residing in urban areas. Among students assigned to treatment group 3, heterogeneous effects were observed among students of Spanish nationality, those living in urban areas, or those with an above-average level of mathematics prior to the intervention. The heterogeneous effects for students in urban areas are particularly notable, with differences ranging from 0.9 to 1.8 points between students receiving treatment (tutorials with 2 students, 3 students, or 2 students plus the socio-emotional component). There are also differences of about one point between students above and below the average level of mathematics (in treatment groups 1 and 3). Finally, the tutoring of 2 students with the socio-emotional component (treatment group 3) has negative effects on the overall grade for students of non-Spanish nationality and students with a level of mathematics prior to the intervention below the average, significant at 5% and 10%, respectively. As for school grades (Table 13.1), the same heterogeneous effects are not observed. Among students assigned to treatment group 1, the school grades of students of Spanish nationality are on average 0.3 points lower than the grades obtained by students who do not have Spanish nationality (significant effect at 10%). It is also found that among students assigned to treatment group 2, the school grades of secondary school students and the grades of students with above-average levels of mathematics prior to the intervention are on average 0.4 and 0.3 points lower than the average grades obtained by primary school students and students below the average level of mathematics before the intervention, respectively (statistically significant effects at 10%). No heterogeneous effects were found among the group of students assigned to treatment group 3.

Table 12.2: Heterogeneous effects on the mathematics level test score in wave 2

	Standardized math test score				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
TG1	0.210	0.140	0.222	0.366*	-0.160
	(0.275)	(0.254)	(0.196)	(0.210)	(2.010)
TG1 x female	0.376				
	(0.376)				
TG1 x Spanish nationality		0.380			
		(0.264)			
TG1 x urban			1.795***		
			(0.504)		
TG1 x secondary school				0.196	
				(0.443)	

	Standardized math test score				
	(1)	(2)	(3)	(4)	(5)
TG1 x prev. grade >50					0.943***
					(0.230)
TG2	0.251	0.225	0.034	0.207	-0.396*
	(0.265)	(0.220)	(0.196)	(0.215)	(0.231)
TG2 x female	-0.167				
	(0.375)				
TG2 x Spanish nationality		-0.092			
		(0.222)			
TG2 x urban			1.337**		
			(0.555)		
TG2 x secondary school				-0.220	
				(0.429)	
TG2 x prev. grade >50					0.328
					(0.244)
TG3	0.023	-0.628**	0.048	0.045	-0.396*
	(0.263)	(0.258)	(0.190)	(0.199)	(0.231)
TG3 x female	0.195				
	(0.365)				
TG3 x Spanish nationality		1.080***			
		(0.260)			
TG3 x urban			0.892		
			(0.574)		
TG3 x secondary				0.491	
				(0.473)	
TG3 x prev. grade >50					0.949***
					(0.244)
Constant	1.485***	1.545***	1.463***	1.468***	1.706***
	(0.211)	(0.213)	(0.210)	(0.212)	(0.218)
Observations	745	745	745	745	745
Control mean	3.005	3.005	3.005	3.005	3.005
TG1 + TG1*X	0.59**	0.52**	2.02***	0.56	0.78***
TG2 + TG2*X	0.08	0.13	1.37***	-0.01	0.31
TG3 + TG3*X	0.22	0.45**	0.94*	0.54	0.55***

Note: Strata-controlled regressions and baseline indicator. Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 13.2: Heterogeneous effects on the school mathematics grade in wave 2

	School grades in mathematics (first quarter of 2023-24)				
	(1)	(2)	(3)	(4)	(5)
TG1	0.223	0.492***	0.292***	0.326***	0.370**
	(0.147)	(0.169)	(0.110)	(0.116)	(9.177)

	School grades in mathematics (first quarter of 2023-24)				
	(1)	(2)	(3)	(4)	(5)
TG1 x female	0.124				
	(0.209)				
TG1 x Spanish nationality		-0.296*			
		(0.173)			
TG1 x urban			-0.038		
			(0.366)		
TG1 x secondary				-0.196	
				(0.265)	
TG1 x prev. grade >50					-0.128
					(0.198)
TG2	0.349**	0.500***	0.292***	0.414***	0.520***
	(0.152)	(0.174)	(0.106)	(0.119)	(0.168)
TG2 x female	-0.054				
	(0.210)				
TG2 x Spanish nationality		-0.271			
		(0.184)			
TG2 x urban			0.352		
			(0.468)		
TG2 x secondary school				-0.435*	
				(0.246)	
TG2 x prev. grade >50					-0.335*
					(0.189)
TG3	0.240*	0.118	0.217**	0.293**	0.241
	(0.142)	(0.177)	(0.110)	(0.115)	(0.164)
TG3 x female	-0.042				
	(0.207)				
TG3 x Spanish nationality		0.136			
		(0.187)			
TG3 x urban			0.040		
			(0.363)		
TG3 x secondary school				-0.384	
				(0.267)	
TG3 x prev. grade >50					-0.024
					(0.186)
Constant	0.482***	0.451***	0.476***	0.487***	0.377**
	(0.114)	(0.117)	(0.113)	(0.113)	(0.147)
Observations	759	759	759	759	759
Control mean	2.734	2.734	2.734	2.734	2.734
TG1 + TG1*X	0.35**	0.20*	0.25	0.13	0.24**
TG2 + TG2*X	0.29**	0.23**	0.64	-0.02	0.19**
TG3 + TG3*X	0.20	0.25**	0.26	-0.09	0.22*

Note: Strata-controlled regressions and baseline indicator. Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Finally, Table 14.2 shows the effect on the self-efficacy index in mathematics for the second wave. Heterogeneous effects were found in all three treatments. In general, among students assigned to the tutorials, students of Spanish nationality, secondary school students, and those with above-average prior knowledge of mathematics have a higher self-efficacy index. The effects are especially large for secondary school students. Compared to the control group, receiving any treatment improves the self-efficacy index among secondary school students by 1.37 to 3.48 standard deviations (significant at 1%). For students above the average level of mathematics before the intervention, the effect of tutoring ranges from 0.45 to 0.65 standard deviations (significant at 5% and 1%, respectively) compared to the control group. These effects are consistent with what was observed in the first wave. Among the groups analyzed, those who benefit the most from tutoring in terms of self-efficacy in mathematics are students enrolled in secondary school and those who had an above-average level of mathematics before starting the tutorials.

Table 14.2: Heterogeneous effects on mathematics self-efficacy in wave 2

	Self-efficacy in mathematics				
	(1)	(2)	(3)	(4)	(5)
TG1	0.347	-0.087	0.346*	0.243	-0.188
	(0.260)	(0.316)	(0.204)	(0.181)	(0.280)
TG1 x female	-0.148				
	(0.359)				
TG1 x Spanish nationality		0.517			
		(0.322)			
TG1 x urban			-0.442		
			(0.345)		
TG1 x secondary school				1.935**	
				(0.799)	
TG1 x prev. grade >50					0.679**
					(0.276)
TG2	0.109	-0.404	0.167	0.141	-0.360
	(0.277)	(0.257)	(0.193)	(0.177)	(0.242)
TG2 x female	0.064				
	(0.358)				
TG2 x Spanish nationality		0.768***			
		(0.268)			
TG2 x urban			-0.082		
			(0.387)		
TG2 x secondary school				1.227**	
				(0.480)	
TG2 x prev. grade >50					0.810***
					(0.251)
TG3	0.399	0.428	0.610***	0.393**	0.139
	(0.268)	(0.297)	(0.200)	(0.179)	(0.251)
TG3 x female	0.155				

	Self-efficacy in mathematics				
	(1)	(2)	(3)	(4)	(5)
	(0.360)				
TG3 x Spanish nationality		0.096			
		(0.309)			
TG3 x urban			-0.806**		
			(0.376)		
TG3 x secondary				3.086***	
				(0.801)	
TG3 x prev. grade >50					0.515*
					(0.265)
Constant	-0.198	-0.217	-0.209	-0.314**	-0.210
	(0.142)	(0.143)	(0.144)	(0.142)	(0.142)
Observations	336	336	336	336	336
Control mean	-0.216	-0.216	-0.216	-0.216	-0.216
TG1 + TG1*X	0.20	0.43**	-0.10	2.18***	0.49***
TG2 + TG2*X	0.17	0.36**	0.08	1.37***	0.45**
TG3 + TG3*X	0.55**	0.52***	-0.20	3.48***	0.65***

Note: Strata-controlled regressions and baseline indicator. Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

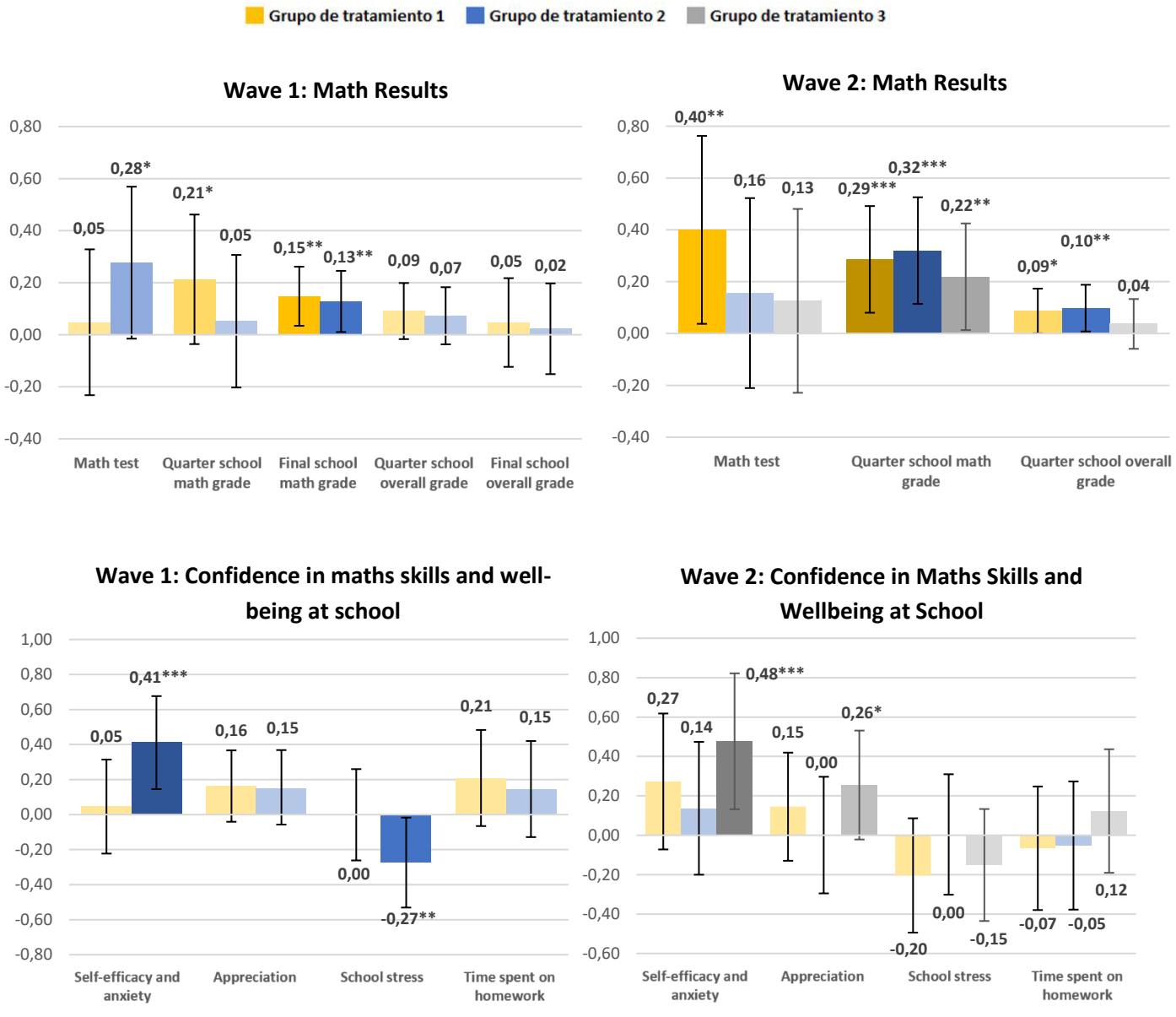
6 Conclusions of the evaluation

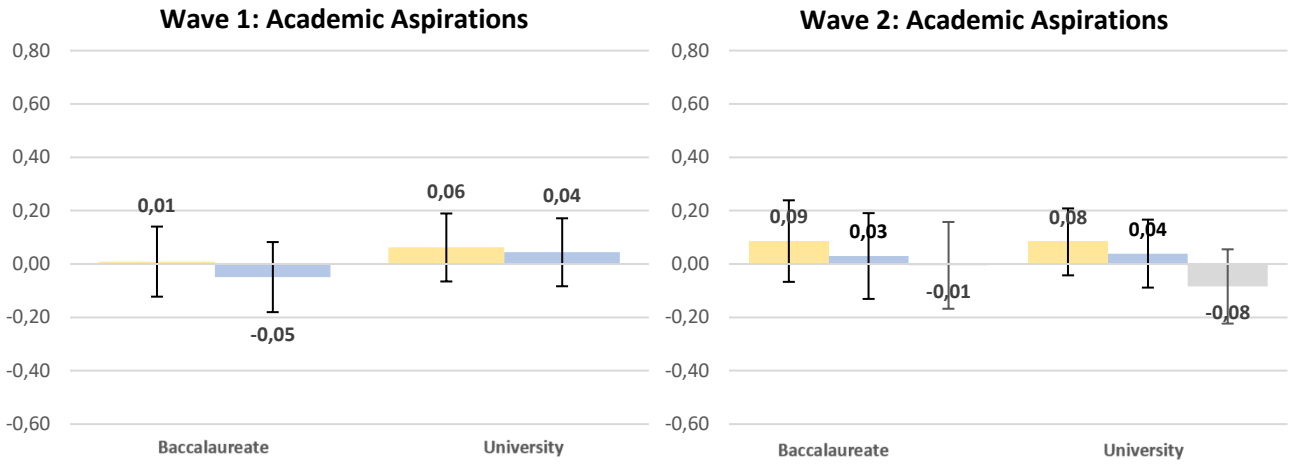
This project has evaluated the effect of offering 8 weeks of online tutoring to students from vulnerable backgrounds, with the aim of improving their academic skills in mathematics. The objective of the evaluation is to test the effectiveness of an intensive online tutoring reinforcement program with 2 students per tutor (treatment group 1), and to measure the differential impact of two alternative treatments: a group of 3 students per tutor (treatment group 2), and a group of 2 students per tutor whose tutor receives additional training in socio-emotional aspects (treatment group 3).

In general, there are positive and statistically significant effects of tutoring on mathematics test outcome indicators, both in standardized level tests and in school grades, the latter being the most robust effect in the different treatment waves and groups. In the first wave, these effects are especially pronounced for students who were assigned to treatment group 2 (tutorials with 3 students), while in the second wave, larger effects are found among students assigned to treatment group 1 (tutorials with 2 students), especially when level test scores are analyzed. Therefore, it can be concluded that tutorials with 3 students do not worsen the results of tutorials in terms of school performance compared to tutorials with 2 students. Regarding the socio-emotional component (included in treatment 3), there are no significant differences with respect to the treatment of tutorials with 2 students. This component refers to the type of training the tutors received prior to the tutorials, not the content of the tutorials themselves, as in all cases the classes were strictly focused on mathematics. On the contrary, some differences were found between the treatment groups when analyzing the index of self-efficacy in mathematics and the index of school stress. The results indicate

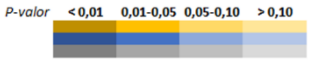
that, in the first wave, tutorials with 3 students (treatment group 2) improved students' confidence in their performance in mathematics and reduced their perception of school stress, while in the second wave, the socio-emotional component in tutors' training generates greater effects on students' confidence in their abilities in mathematics, as well as a greater appreciation for the subject. Regarding academic aspirations, measured with the expectations of studying high school or the desire to study at university, no effect of the online tutoring program for 8 weeks was found.

Figure 7: Effect of the intervention on key indicators





Note: The colors of the figures vary from dark to light depending on the level of significance of the variable's coefficient. The effects included in the graphs refer to the regressions with strata and with baseline indicator.



The heterogeneity analysis indicates that, in general, among the students assigned to receive the tutorials, it is the students who reside in urban centers or with previous grades in mathematics above the average who show the greatest improvements in the results in the mathematics tests. At the same time, there was an increase in self-efficacy and a reduction in anxiety in mathematics among high school students and those with a higher previous level in the subject, regardless of the type of tutoring received. In the second wave, this improvement in the self-efficacy index is also observed among students of Spanish nationality.

In summary, this project captures improvements in academic competencies and in students' confidence in their performance. The results do not show clear evidence in favor of any of the three differential treatments analyzed. Administrative data on tutoring attendance show a high rate of absenteeism, but it is not statistically different between groups. On the other hand, the socio-emotional component of treatment group 3 has a positive effect on variables related to stress and anxiety, but no effect of treatment on other aspects related to the student's socio-emotional well-being is found, and the differences with the rest of the groups are hardly statistically different.

Bibliography

Alegre, F., Moliner, L., Maroto, A., & Lorenzo-Valentin, G. (2019). Peer tutoring in mathematics in primary education: A systematic review. *Educational Review*, *71*(6), 767-791. https://doi.org/10.3386/w27476

Betthäuser, B.A., Bach-Mortensen, A.M. and Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. Nature Human Behaviour, pp.1-11. https://doi.org/10.1038/s41562-022-01506-4

Carlana, Michela; La Ferrara, Eliana (2021). Apart but Connected: Online Tutoring and Student Outcomes during the COVID-19 Pandemic, IZA Discussion Papers, No. 14094, Institute of Labor Economics (IZA). https://www.econstor.eu/bitstream/10419/232846/1/dp14094.pdf

European Commission (2021). Strategic framework for European cooperation in education and training towards the European education area and beyond 2021-2030. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=legissum:4617909

European Commission, European Education and Culture Executive Agency (2023). Promoting diversity and inclusion in schools in Europe, Publications Office of the European Union. https://data.europa.eu/doi/10.2797/443509

COTEC (2023). Meritocracia y Educación: movilidad social y desigualdad de oportunidades. En colaboración con Fundación ISEAK. https://cotec.es/proyectos-cpt/meritocracia-y-educacion/

Decreto Foral 66/2010, de 29 de octubre, por el que se regula la orientación educativa y profesional en los centros educativos de la Comunidad Foral De Navarra. https://www.lexnavarra.navarra.es/detalle.asp?r=9434

Gobierno de España, Ministerio de derechos sociales y agenda 2030 (2022). Plan de Acción Estatal para la Implementación de la Garantía Infantil Europea (2022-2030). Aprobado el 5 de julio del 2022. https://www.mdsocialesa2030.gob.es/derechos-sociales/infancia-y-adolescencia/Garantia_infantil.htm

Gobierno de Navarra (2019). Plan Estratégico de Atención a la Diversidad 2017-2020. Aprobado el 13 de noviembre de 2018. https://gobiernoabierto.navarra.es/es/gobernanza/planes-y-programas-accion-gobierno/plan-estrategico-atencion-diversidad

Gortazar, L., Hupkau, C., & Roldán, A. (2023). Online tutoring works: Experimental evidence from a program with vulnerable children. CEP Discussion Paper No. 1908. https://www.esade.edu/ecpol/en/publications/online-tutoring-works-experimental-evidence-from-a-program-with-vulnerable-children/

Kraft, Matthew A., John A. List, Jeffrey A. Livingston, and Sally Sadoff. (2022). Online Tutoring by College Volunteers: Experimental Evidence from a Pilot Program. AEA Papers and Proceedings, 112: 614-18. https://doi.org/10.1257/pandp.20221038

Ley Orgánica 2/2006, de 3 de mayo, de Educación. https://www.boe.es/eli/es/lo/2006/05/03/2/con

Ley Orgánica 3/2020, de 29 de diciembre, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo, de Educación. https://www.boe.es/eli/es/lo/2020/12/29/3/con

Nickow, A., Oreopoulos, P., & Quan, V. (2020). The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence. https://doi.org/10.1080/00131911.2018.1474176

Q-SER (2020). Proyecto Q6 -El Modelo de Ciclo Circular de Mentoring (4 pasos). https://q-ser.org/es/modelo-de-impacto/

Council Recommendation (EU) 2021/1004 of 14 June 2021 establishing a European Child Guarantee. http://data.europa.eu/eli/reco/2021/1004/oj

Zambrano, V. V., & Gisbert, D. D. (2013). La tutoría entre iguales como un potente recurso de aprendizaje entre alumnos: efectos, fluidez y comprensión lectora. Perspectiva Educacional, Formación de Profesores, 52(2), 154-176. https://www.redalyc.org/pdf/3333/333328170008.pdf

Appendix

Economic and regulatory management

1. Introduction

Within the framework of the National Plan for Recovery, Transformation and Resilience, the General Secretariat of Inclusion of the Ministry of Inclusion, Social Security and Migration is significantly involved in Component 23 "New public policies for a dynamic, resilient and inclusive labor market", framed in policy area VIII "New care economy and employment policies".

Investment 7 "Promotion of Inclusive Growth by linking socio-labor inclusion policies to the Minimum Income Scheme" is one of the reforms and investments proposed in this Component 23. Investment 7 promotes the implementation of a new model of inclusion based on the Minimum Income Scheme (MIS), which reduces income inequality and poverty rates. To achieve this objective, the development of pilot projects has been proposed, among others, for the implementation of social inclusion pathways with the autonomous communities and cities, local entities, and Third Sector of Social Action organizations, as well as with the different social agents.

Royal Decree 938/2021, of October 26, which regulates the direct granting of subsidies from the Ministry of Inclusion, Social Security and Migration in the field of social inclusion, for an amount of €109,787,404, within the framework of the Recovery, Transformation and Resilience Plan[29], contributed to meeting milestone 350 for the first quarter of 2022 as outlined in the Council's Implementing Decision: "Improve the rate of access to the Minimum Income Scheme, and increase the effectiveness of the MIS through inclusion policies, which, according to its description, will translate into supporting the socio-economic inclusion of the beneficiaries of the MIS through itineraries: eight collaboration agreements signed with subnational public administrations, social partners and Third Sector of Social Action to conduct the pathways. The objectives of these partnership agreements are: (i) improve the MIS access rate; ii) increase the effectiveness of the MIS through inclusion policies". Likewise, along with Royal Decree 378/2022, of May 17[30], "at least 10 additional collaboration agreements signed with subnational public administrations, social partners and Third Sector of Social Action organizations to conduct pilot projects to support the socio-economic inclusion of the beneficiaries of MIS through itineraries" contributed to compliance with monitoring

[29] Royal Decree 938/2021, of October 26, 2021, which regulates the direct granting of subsidies from the Ministry of Inclusion, Social Security and Migration in the field of social inclusion, for an amount of 109,787,404 euros, within the framework of the Recovery, Transformation and Resilience Plan (BOE-A-2021-17464). It can be consulted at the following link: https://www.boe.es/diario_boe/txt.php?id=BOE-A-2021-17464.

[30] Royal Decree 378/2022, of May 17, 2022, regulating the direct granting of subsidies from the Ministry of Inclusion, Social Security and Migration in the field of social inclusion, for an amount of 102,036,066 euros, within the framework of the Recovery, Transformation and Resilience Plan (BOE-A-2022-8124). It can be consulted at the following link: https://www.boe.es/diario_boe/txt.php?id=BOE-A-2022-8124.

indicator number 351.1 in the first quarter of 2023, linked to the Operational Arrangements document[31].

In accordance with Article 3 of Royal Decree 378/2022, of May 17, 2022, the granting of subsidies will be carried out by means of a resolution accompanied by an agreement of the head of the Ministry of Inclusion, Social Security, and Migration as the competent authority for granting them, without prejudice to the existing delegations of competence in the matter, upon request of the beneficiary entities.

On **September 22, 2022**, the Chartered Community of Navarre was notified of the Resolution of the General Secretariat for Inclusion and Social Welfare Objectives and Policies granting a subsidy of **2,255,868.00 euros** to the Chartered Community of Navarre and, on **December 13, 2022**, an Agreement is signed between the General State Administration, through the General Secretariat for Inclusion and Social Welfare Objectives and Policies and the Autonomous Community of Navarre for the implementation of a social inclusion project within the framework of the Recovery, Transformation and Resilience Plan, which was published in the "Official State Gazette" on **26 December 2022 (BOE no. 309).**[32]

2. Time frame of the intervention

Article 17(1) of Royal Decree 378/2022 of May 17, 2022, established that the deadline for the implementation of pilot projects for social inclusion itineraries covered by the subsidies provided for in this text shall not exceed the deadline of November 30, 2023, while the evaluation, shall not extend beyond March 31, 2024, in order to meet the milestones set by the Recovery, Transformation, and Resilience Plan with regard to social inclusion policies.

Within this general timeframe, the implementation begins on **March 20, 2023**, with the start of the intervention itinerary, continuing until **November 30, 2023**, and subsequently developing dissemination and evaluation tasks of the project until **March 31, 2024**.

3. Relevant Agents

Among the relevant agents in the implementation of the project are:

- The **Autonomous Community of Navarre**, as the beneficiary entity and coordinator of the project, through:
    a) Department of Education.

[31] Decision of the European Commission approving the document 'Operational Provisions of the Recovery, Transformation and Resilience Plan', which can be consulted at the following link: https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/hacienda/Documents/2021/101121-CountersignedESFirstCopy.pdf.

[32] Resolution of 16 December 2022, of the General Secretariat for Inclusion and Social Welfare Objectives and Policies, which publishes the Agreement with the Autonomous Community of Navarre, for the implementation of a project for social inclusion within the framework of the Recovery, Transformation and Resilience Plan. It can be consulted at the following link: Provision 22460 of BOE no. 309 of 2022

b) Department of Social Rights.

- The **Ministry of Inclusion, Social Security and Migration (MISSM)** as the sponsor of the project, and the main responsible for the RCT evaluation process. The General Secretariat for Inclusion (SGI) assumes the following commitments:
  c) Assist the beneficiary entity in the design of the actions to be carried out for the implementation and monitoring of the object of the grant, as well as for the profiling of the potential participants of the pilot project.
  d) Design the randomized controlled trial (RCT) methodology of the pilot project in coordination with the beneficiary entity.
  e) Evaluate the pilot project in coordination with the beneficiary entity.
- **CEMFI and J-PAL Europe**, as scientific and academic institutions that support MISSM in the design and RCT evaluation.

Sample wear

In the first wave of the experiment carried out in this pilot project, all students responded to the initial survey. However, in the second wave, there was significant attrition in the sample, with only 451 students responding to the survey, while 348 did not respond. To address this issue, a test was conducted to determine whether this lack of response was independent of the assigned treatment group, or if it was related to the assignment of students to the different experimental groups.

Table A-1 shows the difference in participation rates in the second wave's initial survey between the experimental groups. Specifically, it reports the results of linear regressions where the dependent variables are binary indicators equal to one if the learner's data for the initial survey is missing in wave 2. The independent variables are binary indicators equal to one for the assignment to each of the treatment groups, with the control group as the omitted category. It is observed that the higher participation in the initial survey, compared to the control group, is statistically significant at the 10% level for treatment group 1, and at the 1% level for treatment groups 2 and 3. Therefore, it can be concluded that participation rates are statistically different between the control group and the treatment groups.

Table A-1: Participation rate in the initial survey by treatment group for wave 2

	Initial survey (1)
Treatment group 1	0.067* (0.038)
Treatment group 2	0.104*** (0.038)
Treatment group 3	0.130*** (0.038)
Observations	799
Mean control	0.478
Test T1=T2 p.value	0.358
Test T1=T3 p.value	0.114
Test T1=T2=T3 p.value	0.284

Note: Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Sample Balance

Next, Tables 15.1 and 15.2 report the balance tests between the control group and the treatment groups for the first and second waves, respectively. All data in these tables refer to the survey conducted prior to the intervention (baseline) or to time-invariant variables. The mean value of each variable for each group is reported, as well as the number of observations in each group and the p-value resulting from a mean difference test using the student's t-statistic. The lower the p-value, the more confidently one can reject the hypothesis that the mean of the variable in both groups is equal. For example, if the p-value is less than 0.05, the hypothesis of equality of means can be rejected at a 5% confidence level.

Table 15.1: Balance tests between experimental groups for wave 1

Variable	Mean CG N | (var.)	Mean TG1 N | (var.)	Mean TG2 N | (var.)	p-value CG-TG1 N	p-value CG-TG2 N
Age	12.40 164 | (1.36)	12.44 195 | (1.46)	12.54 186 | (1.48)	0.37 359	0.57 350
Spanish nationality	0.67 164 | (0.22)	0.61 195 | (0.24)	0.67 186 | (0.22)	0.20 359	0.96 350
Specific Educational Support Needs Census	0.65 164 | (0.23)	0.66 195 | (0.23)	0.66 186 | (0.23)	0.83 359	0.82 350
Score on Mathematics Level Test	2.95 154 | (2.79)	3.03 190 | (2.55)	3.07 176 | (2.53)	0.88 344	0.84 330
School Grades in Mathematics (1st & 2nd Quarter 2022/23) (0-10)	4.58 163 | (5.35)	4.36 190 | (4.02)	4.44 186 | (4.75)	0.31 353	0.87 349
School Grades in Mathematics (1st & 2nd Quarter 2022/23) (0-5)	1.93 163 | (1.67)	1.69 190 | (1.28)	1.81 186 | (1.42)	0.05** 353	0.65 349
Average School Grade (1st & 2nd Quarter 2022/23) (0-10)	5.68 163 | (2.20)	5.46 191 | (2.13)	5.46 186 | (2.21)	0.10 354	0.40 349
Average School Grade (1st & 2nd Quarter 2022/23) (0-5)	2.33 163 | (1.16)	2.21 191 | (1.12)	2.20 186 | (1.08)	0.13 354	0.62 349
Self-Efficacy and Math Anxiety Index	0.01 122 | (1.02)	0.05 144 | (0.97)	-0.06 139 | (1.02)	0.85 266	0.65 261
Appreciation Level of Spanish/Basque Language	3.73 150 | (1.53)	3.66 185 | (1.10)	3.68 179 | (1.12)	0.86 335	0.91 329
Appreciation Level of Mathematics	3.35 153 | (1.54)	3.18 185 | (1.64)	3.29 181 | (1.53)	0.36 338	0.82 334
Time Spent on Homework	0.06 153 | (1.23)	0.00 182 | (0.87)	-0.05 175 | (0.94)	0.60 335	0.41 328
Plan to Study in High School	0.58 153 | (0.25)	0.58 185 | (0.25)	0.49 181 | (0.25)	0.95 338	0.31 334
Desire to Attend University	0.89	0.88	0.85	0.91	0.45

Variable	Mean CG N | (var.)	Mean TG1 N | (var.)	Mean TG2 N | (var.)	p-value CG-TG1 N	p-value CG-TG2 N
	141 | (0.10)	170 | (0.10)	166 | (0.13)	311	307
School Stress Index	-0.05	-0.05	0.09	0.80	0.31
	142 | (1.26)	174 | (0.73)	169 | (1.05)	316	311
Life Satisfaction Index	-0.00	0.01	-0.00	0.92	0.92
	149 | (1.10)	183 | (1.02)	174 | (0.91)	332	323
Well-Being and School Motivation Index	0.00	0.02	-0.02	0.59	0.71
	139 | (1.04)	159 | (0.92)	166 | (1.05)	298	305
Socio-Emotional Skills Index	0.01	-0.05	0.04	0.84	0.52
	137 | (1.01)	164 | (0.90)	160 | (1.10)	301	297
Locus of Control Index	-0.13	-0.00	0.11	0.45	0.09*
	145 | (1.00)	176 | (0.98)	177 | (1.00)	321	322

Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 15.2: Balance tests between experimental groups for wave 2

Variable	Mean CG N | (var.)	Mean TG1 N | (var.)	Mean TG2 N | (var.)	Mean TG3 N | (var.)	p-value CG-TG1 N	p-value CG-TG2 N	p-value CG-TG3 N
Age	10.78	10.71	10.81	10.78	0.37	0.68	0.36
	209 | (1.50)	198 | (1.54)	204 | (1.56)	188 | (1.72)	407	413	397
Spanish nationality	0.67	0.69	0.68	0.71	0.77	0.50	0.67
	209 | (0.22)	198 | (0.21)	204 | (0.22)	188 | (0.21)	407	413	397
Specific Educational Support Needs Census	0.52	0.54	0.50	0.55	0.28	1.00	0.23
	100 | (0.25)	115 | (0.25)	122 | (0.25)	114 | (0.25)	215	222	214
Score on Mathematics Level Test	2.78	3.14	3.12	2.98	0.17	0.37	0.27
	186 | (2.90)	191 | (2.49)	198 | (3.03)	176 | (3.14)	377	384	362
Grades in Mathematics (3rd Quarter 2022/23) (0-5)	2.77	2.71	2.61	2.51	0.86	0.16	0.21
	208 | (2.30)	198 | (2.20)	203 | (2.10)	187 | (2.17)	406	411	395
Average School Grade (3rd Quarter 2022/23) (0-5)	3.32	3.30	3.26	3.20	0.89	0.40	0.60
	208 | (0.90)	198 | (0.90)	204 | (0.92)	187 | (0.97)	406	412	395
Self-Efficacy and Math Anxiety Index	-0.03	-0.06	0.06	0.03	0.64	0.79	0.76
	85 | (1.13)	87 | (1.19)	97 | (0.84)	92 | (0.89)	172	182	177
Appreciation Level of Spanish/Basque Language	3.74	3.87	3.74	3.83	0.44	0.72	0.74
	100 | (1.27)	115 | (1.01)	119 | (1.13)	114 | (0.95)	215	219	214
Appreciation Level of mathematics	3.65	3.61	3.55	3.72	0.98	0.72	0.74
	100 | (1.66)	115 | (1.71)	121 | (1.52)	114 | (1.44)	215	221	214
School Stress Index	-0.08	0.06	0.13	-0.12	0.44	0.40	0.63
	96 | (0.72)	109 | (0.93)	113 | (1.11)	108 | (1.20)	205	209	204
Life Satisfaction Index	0.06	-0.02	-0.10	0.07	0.59	0.99	0.67
	96 | (0.95)	108 | (1.07)	116 | (1.02)	109 | (0.96)	204	212	205

Variable	Mean CG N | (var.)	Mean TG1 N | (var.)	Mean TG2 N | (var.)	Mean TG3 N | (var.)	p-value CG-TG1 N	p-value CG-TG2 N	p-value CG-TG3 N
Socio-Emotional Skills Index	-0.09	-0.01	0.03	0.06	0.58	0.79	0.08*
	95 | (1.14)	114 | (0.79)	114 | (0.91)	105 | (1.21)	209	209	200
Locus of Control Index	-0.02	-0.03	0.02	0.03	0.57	0.37	0.24
	100 | (1.16)	115 | (1.02)	122 | (0.84)	112 | (1.03)	215	222	212

Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Analysis of results: Results for subcategories and other indicators

Table A-2: Effects on confidence in mathematics and stress in wave 1, subscales

	Mathematics				School Stress		
	Self-Efficacy	Anxiety	Effort	T. interest	Effort	Reward	Overload
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment group 1	0.055	-0.159	0.271**	0.256*	0.082	-0.089	-0.034
	(0.127)	(0.126)	(0.137)	(0.144)	(0.138)	(0.134)	(0.140)
Treatment group 2	0.251*	0.005	0.358**	0.562***	-0.112	-0.172	-0.071
	(0.130)	(0.127)	(0.139)	(0.142)	(0.141)	(0.128)	(0.134)
Observations	307	321	333	340	332	338	342
Mean control	-0.018	-0.061	-0.102	-0.238	-0.050	0.017	0.030
Test T1=T2 p.value	0.092	0.178	0.512	0.015	0.133	0.494	0.771

Note: Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table A-3: Effects on confidence in mathematics and stress in wave 2, subscales

	Mathematics				School Stress		
	Self-Efficacy	Anxiety	Effort	T. interest	Effort	Reward	Overload
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment group 1	0.376**	-0.042	0.066	0.402**	-0.117	-0.057	0.245
	(0.176)	(0.171)	(0.162)	(0.167)	(0.160)	(0.156)	(0.166)
Treatment group 2	0.126	0.003	0.073	0.304	-0.007	0.046	0.300*
	(0.167)	(0.163)	(0.151)	(0.197)	(0.158)	(0.158)	(0.167)
Treatment group 3	0.428**	-0.018	0.172	0.336*	-0.106	-0.041	0.134
	(0.171)	(0.169)	(0.158)	(0.176)	(0.151)	(0.157)	(0.156)
Observations	384	399	404	423	411	404	419
Mean control	-0.224	-0.005	-0.081	-0.235	0.116	0.011	-0.136
Test T1=T2 p.value	0.100	0.727	0.962	0.497	0.418	0.431	0.695
Test T1=T3 p.value	0.724	0.873	0.492	0.622	0.937	0.902	0.409
Test T1=T2=T3 p.value	0.070	0.941	0.747	0.769	0.680	0.709	0.458

Note: Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table A-4: Effects on other well-being and academic interest indicators in wave 1

			School Well-being and Motivation				
	Grit	Locus of control	Scale	Goodwill	Motivation	Life Satisfaction	Appreciation language
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Treatment group 1	-0.056	-0.185	0.003	-0.127	-0.151	0.045	-0.056
	(0.135)	(0.152)	(0.129)	(0.141)	(0.141)	(0.147)	(0.145)
Treatment group 2	0.099	-0.045	0.146	0.027	0.058	0.117	-0.053
	(0.141)	(0.144)	(0.136)	(0.148)	(0.147)	(0.152)	(0.152)
Observations	327	337	320	300	296	330	343
Mean control	-0.033	0.098	0.017	2.180	2.997	-0.072	0.026
Test T1=T2 p.value	0.204	0.275	0.236	0.207	0.117	0.573	0.986

Note: Strata-controlled regressions and baseline indicator. Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table A-5: Effects on other well-being and academic interest indicators in wave 2

	Grit	Locus of control	Life Satisfaction	Appreciation language
Variable	**(1)**	**(2)**	**(3)**	**(4)**
Treatment group 1	-0.017	0.079	0.002	-0.201
	(0.151)	(0.155)	(0.166)	(0.143)
Treatment group 2	-0.106	-0.104	-0.076	-0.151
	(0.153)	(0.159)	(0.173)	(0.156)
Treatment group 3	0.077	0.092	0.195	0.034
	(0.152)	(0.162)	(0.159)	(0.150)
Observations	408	423	415	426
Mean control	-0.090	0.057	-0.025	-0.008
Test T1=T2 p.value	0.459	0.203	0.574	0.706
Test T1=T3 p.value	0.440	0.929	0.102	0.081
Test T1=T2=T3 p.value	0.351	0.334	0.069	0.191

Note: Standard errors were used, reported in parentheses. Levels of significance: * p < 0.10, ** p < 0.05, *** p < 0.01.