Inclusion Policy Lab: Evaluation Results

Fundació Jaume Bofill: educational reinforcement projects - Pentabilities

July 2024

Financiado por la Unión Europea NextGenerationEU
MINISTERIO DE INCLUSIÓN, SEGURIDAD SOCIAL Y MIGRACIONES
Plan de Recuperación, Transformación y Resiliencia
FUNDACIÓ BOFILL Educació per canviar-ho tot
J-PAL

This report has been prepared by the General Secretariat of Inclusion of the Ministry of Inclusion, Social Security, and Migration 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 Fundació Bofill has participated in the writing of this report. This collaborating 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. Likewise, their collaboration has been essential to gathering informed consents, ensuring that participants in the itinerary were adequately informed and that their participation was voluntary.

The following research team has substantially contributed to this study: Caterina Calsamiglia (IPEG), Giacomo de Giorgi (University of Geneva), Laia Navarro-Solà (IIES Stockholm University) and Ece Yagman (IPEG).

The partnership with J-PAL Europe has been a vital component 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 and 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 it 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 affect the interpretation of the results. The researcher is committed to continuing 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 guarantee 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 are 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 projects that reinforce the effectiveness of the MIS: Behavioral evidence for the development of non-cognitive skills - PENTABILITIES", which has been conducted in **cooperation between the MISSM** and the **Fundació Bofill,** an entity of the Third Sector of Social Action, dedicated to promoting and extending new educational opportunities to overcome social inequalities.
- This study consists of a **methodology** focused on the learning of **socio-emotional skills** in Secondary School (ESO in Spain), which aims to test the **causal impact** of implementing formative assessment sessions on social and personal skills on student outcomes.
- The PENTABILITIES program is aimed at schools of **high and maximum** complexity, made up of students in **vulnerable situations**. It is implemented in **Catalonia** and in Andalusia, Ceuta, and Melilla (**Southern Region**).
- For the **treatment group,** the intervention conducted consists of **training** and advising **teaching staff** to integrate the formative assessment of **socio-emotional skills** into the curriculum using digital tools. The **control group** does not receive any specific treatment.
- As for the sample, differentiating between regions, in the case of **Catalonia** the study sample consists of **2,451 students**, aged approximately 12 to 17 years, enrolled from 1st to 4th year of Secondary School (ESO) and **129 teachers from 40 secondary schools**. For its part, in the **southern region** (Andalusia, Ceuta and Melilla), the sample is made up of a total of **45 schools**, **140 teachers** and **2,095 students**.
- Based on **the results of the study**, it is found that the program has had a significant impact on teachers' classroom practices, **improving classroom management** and **strengthening the teacher-student relationship**. Additionally, this study contributes to the debate on how to integrate non-cognitive skills in schools with disadvantaged young people and highlights the role of educational programs in individual development. It offers a promising path for future interventions aimed at reducing gaps in education.

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 responds to 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 Autonomous Communities, the Third Sector of Social Action 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 (onwards, SGI by its acronyms in Spanish) 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 dated December 20, which establishes 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, establishes a framework for the allocation of recovery funds and to boost 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 determine whether 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.

[4] Royal Decree 938/2021, dated October 26, 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 109,787,404 euros, within the framework of the Recovery, Transformation, and Resilience Plan (BOE-A-2021-17464).

[5] Milestone 350 of the RTRP: "Improve the access rate of 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 social action entities of the third sector to conduct the itineraries. These partnership agreements aim to: i) improve the rate of access to the MIS; ii) increase the effectiveness of the MIS through inclusion policies".

[6] RTRP monitoring indicator 351.1: "at least 10 additional collaboration agreements signed with subnational public administrations, social partners and third sector social action entities to conduct pilot projects to support the socio-economic inclusion of MIS beneficiaries through itineraries".

[7] Royal Decree 378/2022, dated May 17, 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).

Additionally, these evaluations provide an exhaustive analysis of the program and its effects, providing insights into why the program was effective, who has benefited most from the interventions, whether there were indirect or unexpected effects, and which components of the intervention worked, and which did 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 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 promoting 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 of Social Action organizations have implemented the programs, collaborating closely in all their facets, including evaluation and monitoring. Additionally, the Ministry has had the academic and scientific support of the Abdul Latif Jameel Poverty Action Lab (J-PAL) Europe and the Center 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 itineraries.

This report refers to “Educational projects that reinforce the effectiveness of the MIS: Behavioral evidence for the development of non-cognitive skills”, executed within the framework of Royal Decree 378/2022[9] by the Fundació Bofill, an entity of the Third Sector of Social Action, dedicated to promoting and extending new educational opportunities to overcome social inequalities. This report contributes to the fulfillment of milestone 351 of the RTRP: 'After the completion of at least 18 pilot projects, 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, dated March 10, which creates the Ethics Committee linked to social inclusion itineraries, on 04/10/2022 issued a favorable report for the implementation of the project that is the subject of the report.

[9] On September 1, 2022, an agreement was signed between the General State Administration, through the SGI and the Jaume Fundació Bofill for the implementation of a project for social inclusion within the framework of the Recovery, Transformation, and Resilience Plan, which was published in the "Boletín Oficial del Estado" on 16 September 2022 (BOE no. 223)

Context of the project

Research in psychology, neuroscience, and education has shown that social and personal skills—such as perseverance, motivation, and teamwork—have a direct impact on increasing educational expectations, as well as on the lives of people and societies. These skills are highly predictive of life achievement and long-term well-being, and a lack of them is linked to a higher likelihood of dropping out of school, violence, addictions, and mental illness (Heckman & Garcia, 2017; Borghans et al., 2008; Kautz et al., 2014; Deming, 2017).

It is widely recognized that, to ensure the adequate development of these skills, it is necessary to incorporate them in an integrated way into the educational learning process. That is, to integrate social and personal skills into the training programs of the curricular subjects, offering a wide range of opportunities where they are worked on and manifested. Additionally, there are two vital periods in which it is most necessary and valuable to promote them: the first years of life (0-3 years) and the period of adolescence (Newman & Dusenbury, 2015). However, the lack of consensus on impartial and reliable measures of these social and personal skills poses a serious challenge to their assessment and development, which is critical for the development of effective educational practices. Therefore, one of the main obstacles to the integration of these skills into the education system has been the absence of a common framework capable of establishing a universal language and understanding.

The objective of the Pentabilities program, as part of the Fundació Bofill project, is to reinforce the effectiveness of the Minimum Income Scheme (MIS) and to evaluate the impact of this socio-educational support program aimed at teachers of students in vulnerable situations, schooled in centers of high and maximum complexity. Pentabilities is a methodology focused on learning socio-emotional skills through formative assessment. This project evaluates the impact of training Secondary School (ESO) teachers to integrate the formative assessment of social and personal skills in the classroom with the help of digital tools on student outcomes.

Regulatory framework associated with the project and the governance structure

In the context of the European Union, member states act autonomously concerning policies and initiatives related to the educational stage. However, this international organization carries out cooperative work to ensure the greatest possible coherence between countries. In this regard, in February 2021, the European Commission published the 2021-2030 Strategy in the field of education and training, which outlines the general objectives to be followed by member states. Specifically, concerning the subject of this report, the objective of achieving a school dropout rate of less than 9% and reducing the percentage of students with low performance in reading, mathematics, and science to 15% stands out.

Likewise, also from the European Commission, Recommendation (EU) 2021/1004 establishing a European Child Guarantee was published in June 2021[10]. This document invites member states to implement a national plan aimed at guaranteeing access to basic health and education rights for children at risk of poverty and social exclusion.

At the international level, projects have also been launched to develop a European assessment protocol for children's social and emotional skills (EAP SEL), to improve students' emotional skills (I-YES), to promote mental health in schools (MH-WB), to empower students to cope with bullying (ENABLE), to develop a resilience curriculum for primary schools in Europe (RESCUR; Cefai et al., 2015), to develop a European Master's Degree in Resilience Education (ENRETE), and to develop teacher training in socio-emotional education competencies such as HOPEs and EMPAQT.

At the national level, the Government of Spain approved the State Action Plan for the Implementation of the European Child Guarantee (2022-2030) in July 2022[11]. Among the challenges identified in this plan, the one that stands out is “promoting educational equity through a comprehensive and flexible education, capable of adapting to individualized needs, especially of the most vulnerable children".

On the other hand, the Education Law[12], 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 conducted. Article 102.1 stands out, which states that lifelong learning is a right and an obligation of all teachers and a responsibility of the education authorities and the schools themselves.

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

[10] 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

[11] 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

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

Considering the educational context, the needs in this aspect, and the potential benefits identified from the reinforcement of social skills, the Fundació Bofill has designed a socio-emotional learning program to enhance social and personal competences in secondary school.

The scientific objective of the project is to explore how a series of observable, measurable, and identifiable non-cognitive skills in educational contexts of active pedagogy can predict and have a causal impact on people's life development (academic, professional, and personal).

The governance framework established for the proper implementation and evaluation of the project includes the following actors:

- The **Fundació Bofill** is the entity responsible for the execution of the project. The Fundació Bofill is a non-profit organization created in 1969 with the aim of promoting transformation and social change in our country. It is a research and proposal laboratory focused on the field of education that works to promote research, debates, and initiatives to generate educational opportunities and combat social inequalities.

- The **Ministry of Inclusion, Social Security, and Migration** (MISSM) is the founding source of the project and responsible for the RCT evaluation. For this reason, the General Secretariat for Inclusion assumes a series of commitments to the Fundació Bofill:
    - Assist the beneficiary entity in the design of the actions to be conducted, for the implementation 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 the scientific collaborators. Likewise, conduct 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, supporting MISSM in the design and RCT evaluation.

Considering all the above, the present report follows this structure: **section 2** provides a **description of the project**, detailing the problems to be solved, the specific interventions, and the target audience to which the intervention is directed. Next, **section 3** contains information related to the **design of the evaluation**, 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**, the analysis of 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, along with a detailed analysis of the econometric analysis conducted and the results for each of the indicators used. The general **conclusions** of the project evaluation are described in **section 6**. Besides, in the **appendix Economic and regulatory**

management, additional information is provided on the management tools and governance of the pilot project.

Ethics Committee linked to the Social Inclusion Itineraries

During research involving human beings 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 the present evaluation on October 4, 2022.

2 Description of the program and its context

This section describes the program that the Fundació Bofill implemented within 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 PENTABILITIES project seeks to improve the conditions of vulnerable students at different stages of secondary education. The project integrates formative assessment of socio-emotional skills into classes of any specialty. The formative assessment to be implemented by the teacher consists of: i) making explicit and sharing the learning objectives in terms of specific behaviors that occur in the classroom, ii) defining moments for the collection of evidence about the student by

oneself, classmates, and the teacher (360º) through innovative digital tools, and iii) providing feedback using the observed results and the definition of new objectives. This type of assessment can impact self-knowledge and make it easier for students to have more tools and resources for personal, academic, and professional life.

The project also seeks to generate evidence regarding:

- The correlation between social skills and academic results.
- The impact of implementing formative assessment sessions on personal skills, as the mechanism that facilitates exogenous variation, on student outcomes.
- The effect of changes in the behaviors analyzed on the student's academic development in the short term.

The project has benefited significantly from the abundant scientific literature available, which has directly influenced its conception and structure. Research in psychology, neuroscience, and education has shown that social and personal skills have a direct impact on the increase in educational expectations, as well as on the lives of people and societies. A deficiency of these is linked to higher school dropouts, violence, addictions, and mental illness.

Although academic success in school partially predicts success in adult life, non-cognitive competencies can predict success in life better than cognitive competencies (Kautz et al., 2014). Concerns that social-emotional education may impair academic learning have proved unfounded; instead, there is clear evidence that social-emotional education helps foster effective learning habits and leads to improved academic performance (Corcoran et al., 2018; Durlak et al., 2011; Taylor et al., 2017). The literature has also shown that socio-emotional training contributes to reducing mental problems in children and young people, such as anxiety, depression, substance abuse, violence, and antisocial behaviors (Barnes et al., 2014; Clarke et al., 2015; Durlak et al., 2011; Korpershoek et al., 2016; OECD, 2015; Sklad et al., 2012; Taylor et al., 2017).

Many successful, evidence-based Social Emotional Learning (SEL) programs have been implemented around the world (Mahoney et al., 2018; Wigelsworth et al., 2016). For example, the Promoting Alternative Thinking Strategies (PATHS) program, implemented in several countries, has demonstrated positive effects on children's aggression rates, social competence, and academic engagement (Greenberg et al., 1995; Shonfeld et al., 2015). In the Canadian context, interventions such as The Roots of Empathy and MindUP have had a positive impact on students' engagement in prosocial behaviors, as well as increasing cognitive control, empathy, and acceptance among their peers (Schonert-Reichl et al., 2012, 2015).

2.2 Target population and territorial scope

The program is implemented in secondary schools. Since the project targets ESO students in vulnerable situations, priority is given to highly complex schools to maximize the number of students from families receiving the MIS.

The project covers the following territories:

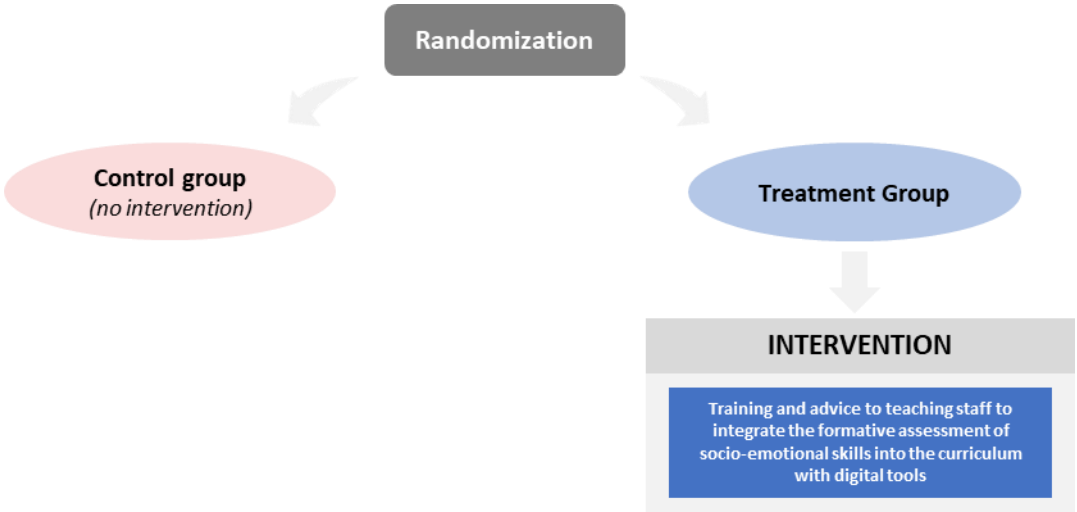
- Autonomous Region of Catalonia.
- Autonomous Region of Andalusia.
- Autonomous Cities of Ceuta and Melilla.

The selection of these regions is based on the Fundació Bofill's capacity to design highly effective and scalable social innovation models, as well as its ability to execute the pilots directly, primarily in Catalonia. Additionally, the foundation's track record enables it to form strategic alliances with key entities to ensure the transfer and implementation of the intervention in other regions. Further details on the recruitment process are provided in Section 3.5, as part of the evaluation design.

2.3 Description of interventions

The proposed intervention consists of training and advising teachers to integrate the formative assessment of social and personal skills in the classroom with the help of digital tools in student outcomes. To rigorously evaluate the impact of the intervention, the participants are distributed in the different courses of each participating center in two different groups, the treatment group, and the control group, which does not receive intervention. **Figure 1** summarizes the interventions that correspond to each group.

Figure 1: Scheme of interventions according to experimental group



For the courses assigned to the treatment group, the intervention consists of training and advising teachers on how to integrate the formative assessment of socio-emotional skills into the curriculum using digital tools. Teachers who taught classes in the courses randomly assigned to the treatment group receive comprehensive training and mentoring on how to implement this assessment in their classrooms. They participate in an initial eight-hour workshop on social and emotional skills, which includes four hours of remote training via videos and four hours of face-to-face training involving case studies and practical group activities. During this workshop, teachers learn how to plan,

observe, record, and provide feedback on their students' social-emotional skills, and become familiar with digital tools for recording and tracking behavioral evidence.

Teachers in the treatment group also have monthly one-on-one meetings with a mentor to discuss potential challenges and solutions related to integrating social-emotional skills development into the classroom. Between four and six classroom observation sessions and mentoring conversations are conducted monthly for each teacher. Additionally, two group sessions are held in which ten to fifteen teachers meet to discuss the implementation of the training received. During these sessions, teachers discuss how to use the evidence gathered to provide feedback to students, as well as the challenges faced and proposed solutions.

Training and mentoring sessions guide teachers in adapting classroom dynamics and planning sessions to promote, observe, and collect evidence on social-emotional skills behaviors within active classroom environments. A time slot is scheduled during sessions to focus on a subset of students, observing and assessing their behaviors. Opportunities are also provided for students to self-evaluate and evaluate their classmates. These assessments are collected regularly for all students using a mobile or desktop app, and the evidence is automatically aggregated over time into individualized reports available on an easy-to-use web platform.

As data accumulates, teachers organize feedback sessions using student reports, highlighting strengths and weaknesses, and comparing evidence from different sources. These sessions are conducted as group reflections or individual meetings with the teacher, resulting in a personalized action plan to further develop the students' socio-emotional skills.

The control group does not receive any specific treatment to avoid 'distorting' the social research and evaluation of the pilot program. Specifically, all centers are treated, with only the group in which the program is implemented varying. Teachers in the control group are offered initial Pentabilities training once the intervention has concluded.

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 understanding 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 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 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 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.

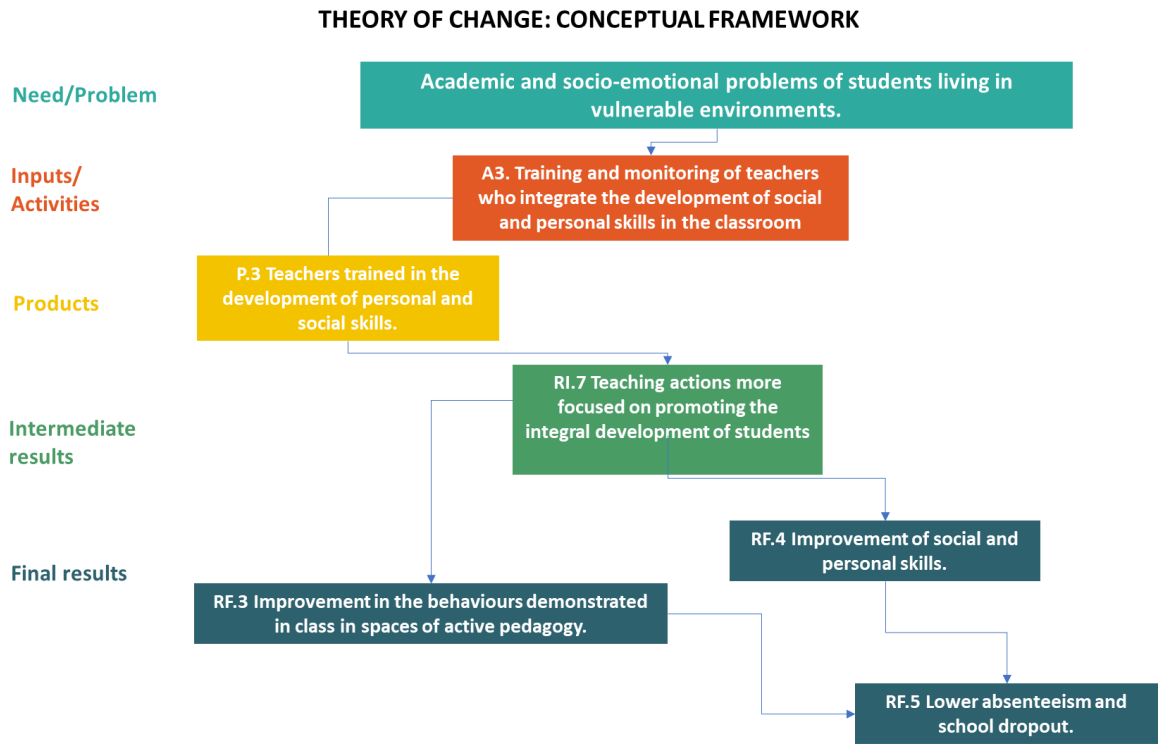
In this context, the Theory of Change serves as a key tool to guide this project, which aims to address the academic and socio-emotional challenges faced by students living in vulnerable environments. Shortcomings in these areas exacerbate inequality and hinder inclusion, as well as equitable access to social opportunities.

This need or problem defines the various areas of action of the project and the activities associated with each. Specifically, the project involves providing training to teachers on personal and social skills, which are expected to significantly contribute to the overall development of their students in a secondary education context characterized by high complexity. The training for the teaching staff includes (1) theoretical-practical sessions on socio-emotional skills and (2) individual and group sessions where constructive feedback will be provided by the mentoring team for the effective implementation of the program.

All these resources and activities produce a series of outputs. By measuring these outputs, it is possible to determine whether the beneficiaries have received the activities or inputs and the extent of their reception. Effective receipt of these resources and activities is crucial for the program to achieve the expected intermediate and results. If beneficiaries do not adequately receive the program, improvements in socio-emotional skills indicators are unlikely. In this project, the output is defined as the number of teachers trained in personal and social skills development.

In the short term, it is anticipated that the training provided to teachers will lead to teaching practices that are more focused on promoting the holistic development of students. This is expected to result in improvements in students' social and personal skills, as well as their behavior in the classroom, contributing to a reduction in absenteeism and school dropout rates.

Figure 2: Theory of change



3.2 Hypothesis

As detailed in the Theory of Change, the ultimate objective of the project is to improve the situation of students at risk of social exclusion through the training and monitoring of teachers who integrate the development of social and personal skills into the classroom. This improvement is reflected in enhanced demonstrated behaviors, improved social and personal skills, and a reduction in absenteeism and school dropout. Consequently, in evaluating the model, various hypotheses are formulated that align with the intermediate and results defined in the Theory of Change. This methodological approach aims to provide a detailed and informed analysis, thus offering a solid basis for strategic decisions in public policy.

The hypotheses to be tested in relation to each of the blocks of results are presented below. The following sections will describe the sources of information for the indicators used in each of the scenarios.

1. Teaching actions more focused on promoting the integral development of students

The hypothesis posits that training and implementing formative assessment of behavior-based socio-emotional skills will change teachers' practices and outcomes in the short term. In addition, the secondary hypotheses propose an increase in teacher motivation, more time spent on active teaching and feedback pedagogy, and a slight average improvement in self-reported socio-emotional skills among teachers who received the intervention.

2. Improved social-emotional skills

The main hypothesis is that behavior-based observational measures of socio-emotional skills, collected by external observers, will be slightly better on average among students receiving the treatment. In turn, the secondary hypothesis posits that self-reported socio-emotional skills will also improve slightly on average among the students receiving the intervention.

3. Improvement in behaviors demonstrated in class in active pedagogy environments

The main hypothesis proposes that self-reported measures of self-awareness regarding one's own socio-emotional skills will be slightly better on average among students who received the treatment. In contrast, the secondary hypothesis suggests that self-reported measures of students' awareness of their peers' socio-emotional skills will improve on average among students who received the treatment.

4. Other Student Outcomes

The main hypothesis is that the formative assessment of behavior-based socio-emotional skills improves other student outcomes in the short term. The secondary hypothesis posits that the well-being and relationship with their primary caregivers is slightly better for the students who received the intervention.

3.3 Sources of information

The responsibility for collecting the information necessary to construct the outcome indicators lies with a team of observers hired by the research team. For data collection, 3 main alternatives are used:

- **Survey**: It consists of questionnaires about the environment and socio-emotional aspects given to the students, which they complete digitally with the assistance of the research team. Specifically, the questionnaires inquire about socio-emotional skills, their level of well-being and the relationship they have with their family and friends. The questionnaires are standardized and internationally validated for the ages of the participants and school settings, such as BESSI, Big5, AWE, Talis, EPOCH, and Autonomous Motivation for Teaching, and have been designed based on scientific literature, including studies such as those of Soto et al. (2022), Soto and John (2017), Waxman et al. (1990), Roth et al. (2007),

Duckworth and Quinn (2009), Baron-Cohen et al. (2001) and Kern et al. (2016). The questionnaires are conducted at different times of the school year: at the beginning (baseline survey), in the middle and at the end of the academic year. Additionally, teachers must also complete questionnaires about their professional and personal context, in addition to assessing their students' socio-emotional skills and their progress during the school year.

The different surveys conducted throughout this report will be referred to as Baseline for the initial survey (pre-data), Midline for the intermediate survey, and Endline for the final survey (post-data).

- **Access to Internal Assessments**: Formative assessment involves the regular collection of indicators that are self-assessed and co-assessed by students, which is conducted weekly or every 15 days, followed by discussions on findings every 6-8 weeks. If the participants give their consent, the evaluation team collects and analyzes this data regularly. Teachers in the treatment group observe, evaluate, and provide *feedback* on their students' social and personal skills. Both teachers and students use the Pentabilities app on devices provided by the schools to record assessments. On the other hand, administrative data from schools is also collected, including records of school performance and achievement, as well as class attendance.
- **Conducting External Observations**: Monthly, a member of the research team collects evidence in person about the educational dynamics and performance of the participants. During these observations, the observer is limited to watching the teacher's lecture without interacting with the students. They also use the Pentabilities app to record students' social and personal skills. They also record the general dynamics of the classroom by applying an active classroom observation methodology to capture the behavior of teachers (TROS). Finally, a standardized activity is conducted in the classrooms to collect measures based on the behavior and socio-emotional skills of the students. This activity consists of building towers in small groups, followed by the evaluation of behaviors by external observers.

3.4 Indicators

This section describes the indicators used for the impact assessment of the pathway, divided by themes related to the scenarios described in **section 3.2**.

Teaching actions more focused on promoting the integral development of students

The following indicators are used to evaluate teaching actions focused on promoting the integral development of students:

Use of the teacher's time: A survey asks teachers to indicate the time spent on each type of activity related to their teaching (both inside and outside the classroom) to assess the impact of the intervention on teaching practices.

Measure of teacher perception (Student's perception of the teacher): Students are asked what their perception of the teacher is.

TROS: This consists of a standardized observation pattern of the teacher's actions in the classroom, collected by external observers. Aspects such as the teacher's interaction with students and the pedagogical strategies used are captured. It is calculated by measuring the percentages of time that teachers are observed in specific instructional variables (Waxman et al., 1990). Higher values indicate a better perception of the teacher in the classroom[13].

BESSI-45 teacher: This is the *Behavioral, Emotional, and Social Skills Inventory, 45-item Short Form* (Soto et al., 2022). It is a composite indicator composed of 5 domains, each with 9 items. Each BESSI proficiency scale is scored by averaging its nine corresponding items.

The 5 domains are: responsibility, cooperation, autonomy and initiative, emotional control, and thinking skills.

BFI-2-S Teacher Score: The *Big Five Inventory – 2 Short Form* (Soto and John, 2017) comprises five domains: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Each domain is scored by summing its 5 facet scales, with responses ranging from 'Strongly Disagree' to 'Strongly Agree'.

Autonomous motivation score to teach: The teacher motivation survey is used in a standardized manner, as described by Roth et al. (2007). To determine an overall motivation score, weights are assigned to participants' scores based on the type of motivation they reflect (–3 for external, –1 for introjection, 1 for identified, and 3 for intrinsic) and then summed.

Growth Mindset Index for Teachers: Measures the belief in the ability to develop skills and competencies through effort and perseverance. Developed by psychologist Carol Dweck (Dweck, C.S., 2000), it contrasts with the fixed mindset. The average rating is calculated by summing the responses to each item and dividing by the total number of items.

Mindset index for teachers. Each item response is coded as either 0 (fixed mindset) or 1 (growth mindset). An average score is calculated from the 7 items, with values ranging between 0 and 1. Adapted from The Education Hub (2020), Mindset Self-Evaluation for Teachers.

Note: *The BFI-2-S is a 30-item version and the BFI-2-XS is an extra-short 15-item version for measuring the big five personality traits.*

For teachers, BESSI-45 is used in the baseline survey and in the endpoint survey and BFI-2-S in both baseline and endline.

[13] Except for item number 8 of the scale ("He criticizes us when we make mistakes"), where the values are inverted.

Improved social-emotional skills

Socio-emotional skills, also referred to as soft skills due to their difficulty in measurement, are assessed using a battery of indices derived from internationally validated questionnaires. Additionally, observations made by external observers will be used during a final activity in which all students participate. Below, we explain the questionnaires and indices used:

Pentabilities 5-Domain Social-Emotional Skills Standardized Score (Autonomy and Initiative, Responsibility, Cooperation, Emotional Management, and Thinking Skills): This score primarily measures social and emotional skills using ratings from external observers on behaviors associated with the target skills. These behaviors are observed in a standardized activity during endline data collection and are combined into 5 indices, one for each Pentabilities domain. The process involves averaging the scores of all observations of a subdomain made by the same observer for a given individual, then calculating the average score at the individual level for each subdomain across all observers. The standardized score for each subdomain is computed at the individual level using the mean and standard deviation of the subdomain in the control group as benchmarks. The average of the standardized subdomain scores at the individual level is then taken and standardized using the benchmarks in the control group. The resulting domain measure is the standardized average of the standardized subdomain scores at the individual level. Fixed effects of observers are included in regressions involving the Pentabilities measurement.

The 5 domains are: Responsibility, Cooperation, Autonomy and Initiative, Emotional Management, and Thinking Skills.

BESSI-20 Score: This is the *Behavioral, Emotional, and Social Skills Inventory, 20-item Short Form* (Soto et al., 2019). Similarly, the BESSI-45 indicator, each proficiency scale of the BESSI-20 is scored by averaging its three or four items. This index measures skills related to the Big Five.

BFI-2-XS score: This is the *Big Five Inventory 2 indicator extra short form (15 items)* (Soto and John, 2017). Each proficiency scale of the BFI-2 is scored by summing its 3 facet scales, which are derived from the 2 corresponding items for each facet.

RMET Index: Reading *the Mind in the Eyes Test* Indicator (Baron-Cohen et al., 2001). Each item is scored as correct or incorrect, and each participant's score is the total number of correct items. This test assesses individuals' empathy.

Perseverance: Again, this is the *Grit-S* indicator (Duckworth and Quinn, 2009). It consists of 8 items with a scale ranging from 5 possible responses, from 'Nothing like me' to 'Very similar to me.' The scores for these 8 items are averaged to create a single index.

Growth mindset index: Measures the belief in the ability to develop skills and competencies through effort and perseverance. Developed by psychologist Carol Dweck, it contrasts with the fixed mindset. The average rating is calculated by summing the responses to each item and dividing by the total number of items.

Note: *The baseline survey for students used the BESSI-45 (45 items) and the BFI-2-S (30 items). Due to time constraints, it was decided to use shorter versions of these questionnaires in the endline survey: the BESSI-20 and the BFI-2-XS. These shorter versions are designed to capture the same latent skill factors, albeit with reduced accuracy. The BFI-2-S is the 30-item version, while the BFI-2-XS is the 15-item extra-short version for measuring the Big Five personality traits.*

Thus, the BESSI-45 is used in the baseline survey and the BESSI-20 in the endline survey for students. The BFI-2-S is used in the baseline survey, and the BFI-2-XS is used in the endline survey for students. For teachers, the BESSI-45 is used in both the baseline and endline surveys, and the BFI-2-S is used in both the baseline and endline surveys.

Improvement in behaviors demonstrated in class in active pedagogy environments

Two indicators are used to verify the improvement in behaviors demonstrated in class within active pedagogy environments:

Self-awareness measure of social-emotional skills level during standardized activity: This measure captures individuals' awareness of their own social-emotional skills using self-reported ratings and is assessed by external observers in a 10-item summative survey collected during the standardized activity.

Measure of Peer Social-Emotional Skills Awareness During Standardized Activity: This measure is constructed by calculating the difference between the student's assessment of a peer's skills and the observer's assessment of the same peer's skills.

Other Student Outcomes

To assess improvements in other student outcomes, two indicators are used:

EPOCH Well-Being Score: This indicator comprises 5 domains: *Engagement, Perseverance, Optimism, Connectedness, and Happiness*. Each item within these domains is scored on a scale from 1 to 5. Scores for each domain are calculated as the average of its items (Kern et al., 2016).

Parental relationship index: This indicator, adapted from Soto et al. (2022), measures the quality of the relationship between students and two of their primary caregivers. For each caregiver, three items are evaluated on a scale from "Very Poor" to "Excellent." The three items for each caregiver are standardized and averaged to form an index.

3.5 Design of the experiment

To evaluate the impact of interventions on the previously mentioned indicators, an experimental evaluation known as a Randomized Controlled Trial (RCT) is used, where participants are randomly assigned to either the treatment group or the control group. The process of recruiting and selecting intervention sites, along with the randomization and timeline of the experiment, is described in detail below.

Recruitment of the centers for the intervention

The process of attracting schools starts with the launch of an information campaign about the project through various media (Twitter, website, and emails to related schools) by the implementing and research team. Additionally, meetings are held with municipalities or centers that may be of interest. In Catalonia, the Department of Education collaborates in the dissemination of the recruitment process through its territorial services.

To apply, schools must complete an email application describing two ESO courses, the number of classes at each level, and between 1-3 teachers per level who could implement the action at each level. The teachers of each level must not overlap to avoid contamination of the control group.

The selection of centers in the program is based on two main criteria: the vulnerability of the target population and methodological limitations. First, indicators of school complexity, the type of school (public, private, or subsidized), and other relevant data, such as the socioeconomic index of the municipality and family income at the census section level, are used to create a representative sample of students and parents with similar socioeconomic and demographic profiles, with an emphasis on vulnerable groups. Second, it is required that each center offer at least two Secondary School courses to ensure a random distribution at different educational levels within each school. To be considered eligible, teachers must select one or more classrooms at a specific grade level (e.g., 1A, 1B, 1C; or just 1A) where they teach at least 3 hours per week in one or more subjects.

Therefore, in the selection process of educational centers, the following aspects are considered:

1. The participating center must have a high level of complexity. Since the study is aimed at vulnerable populations, it is a priority and highly desired characteristic that secondary schools be of high complexity.
2. Courses offered by the school to participate in the project. If a school offers only one course, it has a lower chance of participating compared to a school that offers at least two courses, to ensure that there is both a control course and a treatment course within each center.
3. Whether the school is in a geographical location that is logistically difficult for the implementing or evaluation team to reach.
4. Whether the school is too small with very specific characteristics (e.g., students are not separated by age), as it could be an extreme observation in the treatment or control group.

Once the centers have been defined, they are required to support the collection of different types of informed consents: from the center itself, as well as from families and teachers.

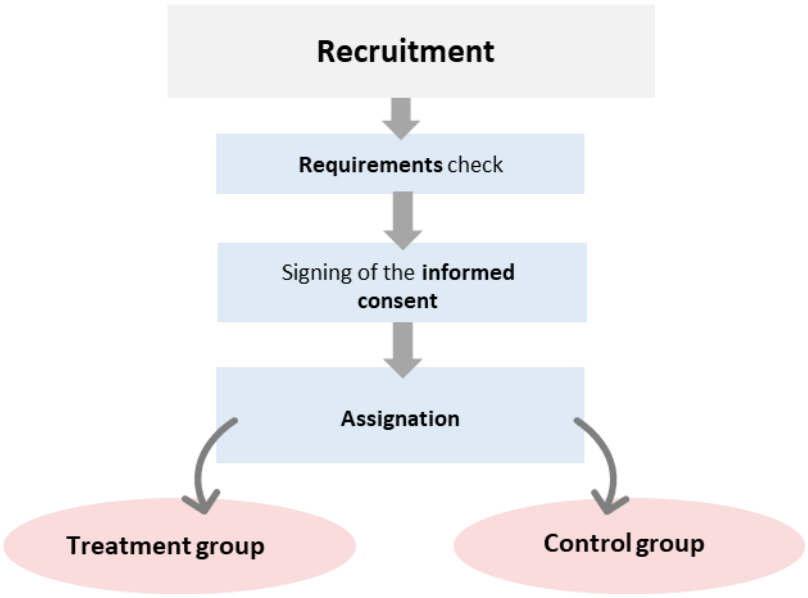
Informed consent

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 harms the validity of the experiment, informed consent may be avoided, or one can give partial information may be given to participants with the approval of the ethics committee.

Random assignment of participants

Figure 3: Sample design

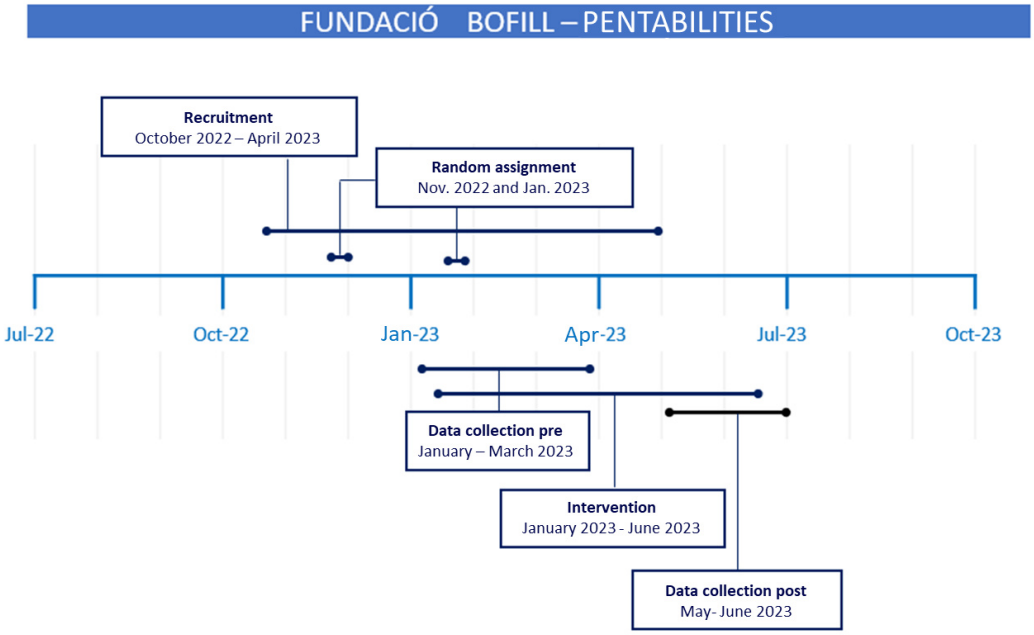


Once the recruitment process is concluded, the process of assigning participants to the treatment and control groups is conducted. The promotions (sets of class groups of the same course) of the participating schools are used as the randomization unit. Randomization is conducted at the promotion level and within each center to determine which course will receive the intervention. In other words, the centers are the randomization strata. Randomization determines which courses are assigned to the treatment group and which to the control group.

This design ensures that each center has both treated and untreated courses, minimizing the risk of overestimating the intervention's effect. If schools only serve as controls, it could discourage families from giving consent to participate and hinder the collection of observational data and quality questionnaires due to lack of motivation. This situation could lead to the identification of an intervention effect, even if it does not exist, due to better attitudes and behaviors in the treated centers for reasons unrelated to the intervention. It is crucial to consider that this design also carries a risk of contamination between courses. Teachers are part of the same faculty, and communication between them could influence the control group's actions. This risk could also affect students who may have siblings or friends in the treated groups. However, this risk is minimized due to the specific nature of the intervention, which involves the application of a very precise methodology that is difficult to implement without access to the application that facilitates systematic data collection and feedback.

Figure 4 shows the time frame in which the implementation and evaluation take place.

Figure 4: Implementation and evaluation timeframe



4 Description of the implementation of the intervention

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

The implementation of the project varied between Catalonia and the South region (Andalusia, Ceuta, and Melilla) in several important aspects. In Catalonia, randomization was conducted at the course or promotion level of Secondary School (ESO), while in the South region it was conducted at the group-class level. Additionally, in the South region, the implementation of the action differed substantially from the implementation in Catalonia. There was no classroom observation, and the intervention started later, limiting the months of intervention and the number of mentoring sessions received. Thus, the expected results are much lower. Consequently, the evaluation of the treatment impact was based solely on surveys, unlike in Catalonia, where external observations and a standardized final activity were also conducted. Due to these differences, the data analysis is presented separately for each region, with tables showing the data for both regions separately, facilitating comparison between them.

4.1 Sample Description

The recruitment strategy focuses on both the municipal level and the school level. Approaches to municipalities and centers are conducted through emails, calls, bulletins, municipal networks, and information sessions.

Of the initial number of potential participants from 303 centers, 99 (32.7%) submitted applications. Of these, 80 schools indicated their willingness and ability to participate by signing the informed consent. This resulted in a potential number of 7,461 participating students. The total number of students contacted was 7,461, of which 5,810 signed their informed consent. Table 1a and Table 1b summarize the main results of the recruitment process, with a breakdown of the different programs and territories where they were conducted.

Table 1a: Record of the recruitment process (centers)

Territorial scope	Catalonia	Andalusia	Ceuta and Melilla	Total
Potential participating centers	**259**	**51**	**13**	**303**
Number of centers contacted	**54**	**51**	**13**	**99**
Number of centers that sign the IC	**39**	**37**	**7**	**80**

Table 1b: Record of the recruitment process (students)

Territorial scope	Catalonia	Andalusia, Ceuta, and Melilla	Total
Potential participating students	4,568	2,893	**7,461**
Number of students signing the IC	**2,905**	**2,095**	**5,810**

Among the reasons for centers refusing to participate in the project, some did not meet the requirements, while others declined due to the high involvement required from teachers and students.

In Catalonia, the study included 39 secondary schools, with the consent of 2,905 students and 184 teachers. Due to the high resource needs for data collection and classroom observation, a sample of 2,451 students and 129 teachers was chosen, comprising those students who received more hours from the participating teachers. In southern Spain, the intervention was implemented in 45 schools, with the initial participation of 2,095 students and 140 teachers who gave their consent. From these participants, a sample of 1,866 students and 102 teachers was selected for the survey.

Final Assessment Sample Characteristics

Table 2a presents the main descriptive statistics of the study sample in Catalonia. A gender balance is observed, with 52% of students being male and 48% female. Regarding the distribution by year of Secondary School (ESO), 74% of students are in the first cycle (42% in 1 ESO and 32% in 2 ESO), while the remaining 26% are in the second cycle (22% in 3 ESO and 4% in 4 ESO). Consequently, the average age of students is 13.3 years.

Regarding the country of birth, 87% of the students in the sample were born in Spain, while the remaining 13% come from other countries, mainly Morocco, Pakistan, and Peru. Considering the country of birth of the main caregiver, 39% of students have parents born outside Spain.

Geographically, the sample is mostly concentrated in Barcelona (88% of students), with representation from the regions of Lleida (8%), Girona (2%), and Tarragona (2%). As explained above, in the selection of educational centers, those with a high degree of complexity were prioritized. As a result, 28% of the students in the sample belong to medium-high complexity centers, 31% to high complexity centers, and 9% to very high complexity centers. The remaining 33% come from medium-low complexity centers.

Table 2a: Descriptive statistics of the sample (Catalonia)

Variable	N	Mean	Standard deviation	Minimum	Maximum
Student Characteristics					
Female	2,164	0.48	0.50	0	1
Age in 2022	1,938	13.31	1.03	12	17
Country of Birth: Spain	2,164	0.87	0.34	0	1
Country of Birth: Other	2,164	0.13	0.34	0	1
Country of Birth: Morocco	2,412	0.01	0.12	0	1
Country of Birth: Pakistan	2,412	0.01	0.12	0	1
Country of Birth: Peru	2,412	0.01	0.09	0	1
Residences: Barcelona	2,412	0.88	0.33	0	1
Residences: Lleida	2,412	0.08	0.27	0	1
Residences: Girona	2,412	0.02	0.15	0	1
Residences: Tarragona	2,412	0.02	0.14	0	1

Variable	N	Mean	Standard deviation	Minimum	Maximum
School Features					
Complexity Index: Medium-Low	2,412	0.33	0.47	0	1
Complexity Index: High	2,412	0.31	0.46	0	1
Complexity Index: Medium-High	2,412	0.28	0.45	0	1
Complexity Index: Very High	2,412	0.09	0.28	0	1
Characteristics of the ESO Level and Class					
1st Course	2,412	0.42	0.49	0	1
2nd Course	2,412	0.32	0.47	0	1
3rd Course	2,412	0.22	0.42	0	1
4th Course	2,412	0.04	0.19	0	1
Number of hours per class	2,412	7.68	5.40	2	35
Number of subjects per class	2,412	2.82	1.51	1	9
Characteristics of the Primary Caregiver					
Country of Birth: Spain	2,143	0.61	0.49	0	1
Country of Birth: Other	2,143	0.39	0.49	0	1
Country of Birth: Morocco	2,143	0.09	0.29	0	1
Country of Birth: Peru	2,143	0.02	0.14	0	1
Country of Birth: Pakistan	2,143	0.02	0.13	0	1
Education: I do not know	1,265	0.46	0.50	0	1
Secondary education	1,265	0.22	0.41	0	1
Vocational training	1,265	0.21	0.41	0	1
Primary education	1,265	0.08	0.27	0	1
Education: None	1,265	0.04	0.20	0	1
People at home	1,265	3.12	1.00	1	5
Rooms at home	1,265	3.15	0.86	1	6

Table 2a (continued): Descriptive statistics of the sample (Catalonia)

Variable	N	Mean	Standard deviation	Minimum	Maximum
Use of teacher time					
Direct instruction to the whole class (Master classes)	105	1.99	1.03	1	5
Working with students individually	104	1.95	0.99	1	5
Lead the class in small groups	104	2.04	1.03	1	5
Addressing student disciplinary issues	104	1.62	0.75	1	5

Variable	N	Mean	Standard deviation	Minimum	Maximum
Facilitate group work	105	2.29	1.05	1	5
Addressing student disciplinary issues	103	1.81	0.82	1	5
Meetings with the rest of the teaching staff/management	104	2.4	0.73	1	5
Delivery of evaluations	101	1.8	0.93	1	5
Use assessment results	103	1.65	0.74	1	4
Supervisory features	105	1.86	0.78	1	5
Professional development activities	104	2.31	1.03	1	5
Preparing for state assessments, etc.	101	1.56	0.91	1	5
Individual planning time	105	3.44	1.07	1	5
Communicate with the students' parental figures	104	3.02	1.17	1	5
Collaborative planning time	103	3.03	1.02	1	5
Administrative duties	103	2.78	1.23	1	5
Pre-Instruction	105	3.36	1.03	1	5
Post-instruction	103	3.42	1.08	1	5
Student's perception of the teacher (11-items)					
It makes me feel like I matter	2,140	3.61	1.13	1	5
Show interest in my work	2,141	4.02	0.96	1	5
Motivates me to succeed	2,139	3.88	1.12	1	5
It encourages me to explain my thought process	2,143	3.74	1.13	1	5
Involve the whole class	2,143	4.17	0.96	1	5
Praises us for good ideas	2,141	4.11	1.02	1	5
Encourages us to ask/comment in lessons	2,142	4.11	0.98	1	5
He criticizes us when we make mistakes	2,140	1.87	1.17	1	5
Encourages student self-management	2,134	3.84	0.92	1	5
Facilitates conflict resolution	2,139	3.88	1.05	1	5
Encourages us to help each other	2,143	4.05	0.99	1	5
TROS Descriptives					
Environment					
Whole-class instruction (%)	112	0.5	0.3	0	1
Small Group Instruction (%)	112	0.21	0.26	0	1
Individual (%)	112	0.25	0.26	0	1

Variable	N	Mean	Standard deviation	Minimum	Maximum
Other (%)	112	0.04	0.09	0	1
Interaction					
No interaction (%)	112	0.11	0.14	0	0.7
Student/Instructional (%)	112	0.59	0.28	0	1
With student/manager (%)	112	0.2	0.19	0	0.78
With student/social (%)	112	0.04	0.07	0	0.25
Student-Based/Collaborative (%)	112	0.05	0.12	0	0.6
With student/others (%)	112	0.01	0.04	0	0.3
Nature of the interaction					
Questioning (process) (%)	112	0.06	0.1	0	0.4
Questioning (content) (%)	112	0.15	0.2	0	0.88
Explanation (%)	112	0.41	0.21	0	1
Comment (%)	112	0.07	0.17	0	0.67
Listen (%)	112	0.08	0.11	0	1
Indications or suggestions (%)	112	0.06	0.11	0	0.4
Modeling/Demonstration (%)	112	0.08	0.15	0	0.62
Other (%)	112	0.1	0.13	0	0.56
Purpose of the interaction					
Content Focus (%)	127	0.4	0.29	0	1
Process Focus (%)	127	0.26	0.23	0	0.9
Product Focus (%)	127	0.18	0.19	0	0.88
Connect content to other disciplines (%)	127	0.01	0.03	0	0.2
Present multiple perspectives on the topic (%)	127	0.05	0.12	0	0.67
Redirect the student's thinking (%)	127	0.08	0.14	0	0.57
Show interest in the student's work (%)	127	0.15	0.17	0	0.71
Show personal appreciation for the student (%)	127	0.05	0.1	0	0.5
Encourage mutual aid between students (%)	127	0.01	0.05	0	0.25
Motivate students to succeed (%)	127	0.11	0.17	0	0.7
Encourage students to question (%)	127	0.02	0.05	0	0.33
Encourage extensive student responses (%)	127	0.06	0.12	0	0.56
Encourage student self-management (%)	127	0.08	0.15	0	0.71

Variable	N	Mean	Standard deviation	Minimum	Maximum
Praise student behavior (%)	127	0.01	0.04	0	0.17
Praise student performance (%)	127	0.04	0.09	0	0.5
Correct student behavior (%)	127	0.09	0.15	0	1
Other (%)	127	0.08	0.13	0	0.62
Bessi-45 professors					
Self-management	105	3.64	0.53	2	5
Social Commitment	105	3.41	0.48	2	5
Cooperation	105	3.92	0.49	3	5
Emotional Management	105	3.49	0.6	2	5
Innovation	105	3.52	0.53	2	5
BFI-2-S Teacher Index					
Extroversion	105	3.82	0.61	2	5
Amiability	105	4.12	0.5	3	5
Responsibility	105	3.97	0.81	2	5
Negative Emotionality	105	2.48	0.67	2	4
Open-mindedness	105	3.97	0.75	2	5
Autonomous motivation to teach					
External Motivation	105	2.47	0.86	1	4
Unconscious Motivation	105	3.71	0.85	1	5
Motivation Identified	105	4.73	0.41	2	5
Intrinsic motivation	105	4.62	0.44	2	5
*Growth mindset**					
DMI Index	111	2.41	1.07	1	5
*Mindset for teachers**					
Teacher Mindset Index	113	4.37	0.31	3	5
Bessi-45 students					
Self-management	2,181	3.44	0.62	1	5
Social Commitment	2,181	3.23	0.67	1	5
Cooperation	2,181	3.61	0.57	1	5
Emotional Management	2,181	3.14	0.77	1	5
Innovation	2,181	3.28	0.62	1	5
BFI-2-S Student Index					
Extroversion	2,177	3.38	0.67	1	5

Amiability	2,176	3.77	0.64	1	5
Responsibility	2,176	3.34	0.68	1	5
Negative Emotionality	2,176	2.84	0.74	1	5
Open-mindedness	2,176	3.44	0.68	1	5
RMET Index					
RMET % Score	2,170	0.67	0.12	0	1
EPOCH					
Commitment	2,157	3.14	0.85	1	5
Perseverance	2,157	3.59	0.86	1	5
Optimism	2,157	3.28	0.96	1	5
Connectivity	2,157	3.99	0.92	1	5
Happiness	2,157	3.71	0.98	1	5
Parental Relationship Index					
Caregiver Index 1	2,138	4.24	0.48	1	6
Caregiver Index 2	2,054	4.15	0.5	1	6
*Student Perseverance Index**					
Grit Index,	1,148	3.34	0.59	1	5
*Growth Mindset Index (DMI)**					
DMI Index	1,131	3.76	1.18	1	6
Pentabilities 5 Social-Emotional Skill Domains Score					
Autonomy	1,868	0.01	0.98	-2	2
Cooperation	1,871	0.02	0.99	-3	3
Emotional management	1,855	0.02	0.99	-3	3
Responsibility	1,858	0.02	1	-3	2
Thought	1,855	-0.02	1.01	-2	3
*Self-reported Bessi-20**					
Self-management	2,163	3.53	0.68	1	5
Social Commitment	2,163	3.21	0.77	1	5
Cooperation	2,163	3.67	0.67	1	5
Emotional Management	2,163	3.22	0.8	1	5
Innovation	2,163	3.35	0.71	1	5
*Bessi-20 observed by teachers**					
Self-management	1,316	3.23	1.03	1	5
Social Commitment	1,315	3.12	0.89	1	5
Cooperation	1,315	3.35	0.79	1	5
Emotional Management	1,315	3.32	0.82	1	5

Innovation	1,316	3.27	0.81	1	5

Note: Measures with * indicate that outcomes were only measured at the endline.

As **Table 2b** shows, the project in southern Spain region presents a situation quite similar to that described for Catalonia.

The sample is mainly concentrated in Andalusia, representing 87% of the total and distributed among the different provinces of the region as follows: Huelva (22%), Málaga (19%), Jaén (16%), Almería (12%), Cádiz (7%), Seville (6%), and Granada (4%). The remaining participants reside in Ceuta (11%) and Melilla (2%). Regarding the characteristics of the students, the gender ratio is equitable, and most students (96%) were born in Spain. Additionally, the data show that 33% of students are in 1st year of ESO, 32% in 2nd ESO, 20% in 3rd ESO, and 15% in 4th ESO.

Regarding the characteristics of the main caregiver, there is a marked predominance of those born in Spain (84%) compared to other countries. In terms of education, less than half (49%) have attended university. Additionally, on average, households have about five people and between three and four bedrooms.

For brevity, information on the results is not included in the descriptive table for the Southern Region.

Table 2b: Descriptive statistics of the sample (Southern Region)

Variable	N	Mean	Standard deviation	Minimum	Maximum
Student Characteristics					
Female	1,598	0.48	0.50	0	1
Age in 2022	1,593	13.80	1.23	12	17
Country of Birth: Spain	1,737	0.96	0.20	0	1
Country of Birth: Other	1,737	0.04	0.20	0	1
Country of Birth: Other: Morocco	1,878	0.01	0.12	0	1
Country of Birth: Other: Colombia	1,878	0.01	0.07	0	1
Country of Birth: Other: Venezuela	1,878	0.00	0.06	0	1
Country of Birth: Other: Other	1,878	0.98	0.15	0	1
Residences: Huelva	1,878	0.22	0.42	0	1
Residences: Malaga	1,878	0.19	0.39	0	1
Residences: Jaén	1,878	0.16	0.37	0	1
Residences: Almería	1,878	0.12	0.33	0	1
Residences: Ceuta	1,878	0.11	0.31	0	1

Variable	N	Mean	Standard deviation	Minimum	Maximum
Residences: Cádiz	1,878	0.07	0.26	0	1
Residences: Seville	1,878	0.06	0.23	0	1
Residences: Granada	1,878	0.04	0.20	0	1
Residences: Melilla	1,878	0.02	0.15	0	1
Characteristics of the Degree					
1st ESO	1,878	0.33	0.47	0	1
2nd ESO	1,878	0.32	0.47	0	1
3rd ESO	1,878	0.20	0.40	0	1
4th ESO	1,878	0.15	0.36	0	1
Characteristics of the Primary Caregiver					
Country of Birth: Spain	1,720	0.84	0.37	0	1
Country of Birth: Other	1,720	0.16	0.37	0	1
Country of Birth: Other: Morocco	1,878	0.08	0.28	0	1
Country of Birth: Other: Colombia	1,878	0.01	0.08	0	1
Country of Birth: Other: Argentina	1,878	0.01	0.08	0	1
Country of Birth: Other: Other	1,878	0.90	0.30	0	1
Education: Attend College	1,718	0.49	0.50	0	1
Education: Did not attend college	1,718	0.34	0.47	0	1
Education Did not attend college. Primary	1,142	0.12	0.33	0	1
Education Did not attend college. High school	1,142	0.24	0.43	0	1
Education Did not attend college. FP	1,142	0.24	0.42	0	1
Education Did not attend college. I do not know	1,142	0.36	0.48	0	1
Education Did not attend college. No	1,142	0.04	0.20	0	1
Education: I do not know	1,142	0.36	0.48	0	1
People in the house	1,605	5.19	9.77	1	99
Bedrooms in the house	1,605	3.53	5.93	1	99

4.2 Random Assignment Results

Once the participating centers were selected, they were randomly assigned to the control group or the treatment group, as explained in **section 3.5. Table 3** presents the results of the randomization process.

Table 3. Random Assignment Results

	Population			Sample		
	Total	TG	CG	Total	TG	CG
Student Level						
Catalonia	2,905	1,401	1,504	2,451	1,244	1,207
Southern Spain	2,095	1,088	1,007	1,748	885	863
Teacher Level						
Catalonia	184	95	89	129	62	67
Southern Spain	140	71	69	121	63	58

Note: The population described includes all students and faculty with consent to participate in the study. The potential sample is the group of teachers and students who, for logistical reasons, we seek to survey.

To verify that the random assignment defines statistically comparable control and treatment groups, an equilibrium test is conducted to ensure that, on average, the observable characteristics of participants in both groups are the same. The balance between the experimental groups is key to inferring the causal effect of the program by comparing its results.

As shown in Table 4, in Catalonia, the sample is balanced in terms of sociodemographic characteristics. For the characteristics described, the mean difference between the treated group and the control group is not statistically significant in any case. Additionally, the number of observations is almost identical in both groups. These two facts confirm that the randomization was conducted correctly.

Table 4: Equilibrium tests between experimental groups (Catalonia)

	Control Group		Treatment Group		Total		
	N	Mean	N	Mean	N	Dif. Of means	p-value
Student Characteristics							
Male	547	0.51	548	0.51	1,095	0.005	0.687
Female	461	0.47	468	0.48	929	0.009	0.791
Age in 2022	973	13.26	969	13.38	1,942	0.114**	0.021
Spain	864	0.89	840	0.87	1,704	-0.022	0.165
Home Features							
People in the house	985	3.10	981	3.14	1,966	0.037	0.412
Bedrooms in the house	984	3.18	980	3.12	1,964	-0.062	0.112
Country of Birth of Caregivers							
Spain	654	0.61	648	0.61	1,302	-0.004	0.886
School Province							
Barcelona	1,058	0.89	1,057	0.87	2,115	-0.020	0.167

	Control Group		Treatment Group		Total		
	N	Mean	N	Mean	N	Dif. Of means	p-value
Lleida	90	0.08	101	0.08	191	0.007	0.850
Girona	29	0.02	30	0.02	59	0.000	0.994
Tarragona	16	0.01	31	0.03	47	0.012	0.789

Note: The sample includes all randomized students with baseline and endpoint levels. Layer effects are included for the level of complexity. Standard errors are grouped at the school-grade level. * P<0.10, **P<0.05, ***P<0.01

Table 4 (continued):

Equilibrium tests between experimental groups (Catalonia)

	Control Group		Treatment Group		Total		
	N	Mean	N	Mean	N	Dif. Of means	p-value
Use of teacher time							
Direct instruction to the whole class (Master Classes)	51	2.04	52	1.92	103	-0.088	0.566
Working with students individually	51	2.02	51	1.90	102	-0.107	0.553
Lead the class in small groups	50	2.02	52	2.08	102	0.053	0.783
Addressing student disciplinary issues	51	1.61	51	1.63	102	0.002	0.897
Facilitate group work	51	2.33	52	2.23	103	-0.134	0.620
Addressing student disciplinary issues	51	1.82	50	1.78	101	-0.049	0.792
Meetings with the rest of the teaching staff/management	51	2.35	51	2.45	102	0.127	0.503
Delivery of evaluations	49	1.84	50	1.80	99	-0.068	0.845
Use assessment results	51	1.65	50	1.66	101	-0.029	0.931
Supervisory features	51	1.86	52	1.87	103	-0.000	0.986
Professional development activities	51	2.25	51	2.35	102	0.082	0.631
Preparation for state assessments, etc.	50	1.60	49	1.55	99	-0.067	0.792
Individual planning time	51	3.71	52	3.17	103	-0.550**	0.012
Communicate with the students' parental figures	51	3.27	51	2.75	102	-0.503**	0.022
Collaborative planning time	51	3.27	50	2.80	101	-0.460**	0.017
Administrative duties	51	2.65	50	2.90	101	0.219	0.308
Pre-Instruction	51	3.59	52	3.15	103	-0.434**	0.031

	Control Group		Treatment Group		Total		
	N	Mean	N	Mean	N	Dif. Of means	p-value
Post-instruction	51	3.59	50	3.28	101	-0.319	0.146
Student's perception of the teacher (11-items)							
It makes me feel like I matter	1,058	3.55	1,082	3.67	2,140	0.112**	0.021
Show interest in my work	1,058	3.98	1,083	4.06	2,141	0.079*	0.058
Motivates me to succeed	1,057	3.84	1,082	3.93	2,139	0.092*	0.056
It encourages me to explain my process...	1,059	3.70	1,084	3.78	2,143	0.081*	0.100
Involve the whole class	1,059	4.10	1,084	4.25	2,143	0.151***	0.000
Praises us for good ideas	1,058	4.01	1,083	4.21	2,141	0.206***	0.000
Encourage us to ask/comment...	1,058	4.04	1,084	4.19	2,142	0.148***	0.000
He criticizes us when we make mistakes	1,058	1.86	1,082	1.88	2,140	0.024	0.641
Encourages student self-management	1,056	3.83	1,078	3.84	2,134	0.007	0.857
Facilitates conflict resolution	1,057	3.86	1,082	3.90	2,139	0.045	0.325
Encourages us to help each other	1,059	4.00	1,084	4.10	2,143	0.100**	0.020

Note: The sample includes all randomized students with baseline and endpoint levels. * p<0.10, **p< 0.05, ***p<0.01

Table 4 (continued):

Equilibrium tests between experimental groups (Catalonia)

	Control Group		Treatment Group		Total		
	N	Mean	N	Mean	N	Dif. Of means	p-value
TROS							
Environment							
Whole-class instruction (%)	54	0.45	58	0.54	112	0.077	0.114
Small Group Instruction (%)	54	0.21	58	0.22	112	0.012	0.827
Individual (%)	54	0.30	58	0.20	112	-0.095**	0.034
Other (%)	54	0.04	58	0.04	112	0.006	0.886
Interaction							
No interaction (%)	54	0.15	58	0.07	112	-0.083***	0.001

	Control Group		Treatment Group		Total		
	N	Mean	N	Mean	N	Dif. Of means	p-value
Student/instructional (%)	54	0.57	58	0.62	112	0.048	0.344
With student/managerial (%)	54	0.19	58	0.21	112	0.017	0.569
With student / social (%)	54	0.04	58	0.04	112	-0.007	0.688
With student / collaborative (%)	54	0.04	58	0.06	112	0.021	0.457
With student/others (%)	54	0.01	58	0.01	112	0.003	0.656
Nature of the Interaction							
Questioning (process) (%)	54	0.08	58	0.04	112	-0.040**	0.050
Questioning (content) (%)	54	0.12	58	0.17	112	0.048	0.205
Explanation (%)	54	0.39	58	0.43	112	0.035	0.306
Comment (e.g., general discussion) (%)	54	0.08	58	0.06	112	-0.022	0.505
Listen (%)	54	0.08	58	0.08	112	0.006	0.751
Indications or suggestions (%)	54	0.05	58	0.07	112	0.022	0.303
Modeling/Demonstration (%)	54	0.10	58	0.05	112	-0.046*	0.064
Other (%)	54	0.10	58	0.10	112	-0.004	0.818
Purpose of the interaction							
Content Focus (%)	54	0.40	58	0.45	112	0.046	0.336
Process Focus (%)	54	0.27	58	0.28	112	0.004	0.914
Focus on the product (e.g., outcome) (%)	54	0.23	58	0.16	112	-0.067	0.103
Connect content to other disciplines (e.g., outcome) (%)	54	0.01	58	0.01	112	-0.002	0.806
Present multiple perspectives on the topic (%)	54	0.03	58	0.07	112	0.031	0.117
Redirect the student's thinking (%)	54	0.09	58	0.10	112	0.022	0.502
Show interest in the student's work (%)	54	0.11	58	0.15	112	0.035	0.301
Show personal appreciation for the student (%)	54	0.06	58	0.08	112	0.015	0.547
Encourage mutual aid between students (%)	54	0.02	58	0.01	112	-0.007	0.390
Motivate students to succeed (%)	54	0.08	58	0.11	112	0.038	0.160
Encourage students to question (%)	54	0.04	58	0.03	112	-0.013	0.567
Encourage extensive student responses (%)	54	0.03	58	0.03	112	-0.001	0.897

	Control Group		Treatment Group		Total		
	N	Mean	N	Mean	N	Dif. Of means	p-value
Encourage student self-management (%)	54	0.04	58	0.03	112	-0.012	0.446
Praise student behavior (%)	54	0.02	58	0.01	112	-0.013	0.259
Praise student performance (%)	54	0.02	58	0.05	112	0.026*	0.083
Correct student behavior (%)	54	0.08	58	0.11	112	0.025	0.362
Other (%)	54	0.08	58	0.07	112	-0.012	0.494

Note: The sample includes all randomized students with baseline and endpoint levels. * p<0.10, **p< 0.05, ***p<0.01

Table 4 (continued):

Equilibrium tests between experimental groups (Catalonia)

	Control Group		Treatment Group		Total		
	N	Mean	N	Mean	N	Dif. Of means	p-value
BESSI-45 professors							
Self-management	52	3.77	53	3.52	105	-0.256**	0,013
Social Commitment	52	3.41	53	3.42	105	0.009	0,923
Cooperation	52	3.99	53	3.84	105	-0.148	0,120
Emotional Management	52	3.54	53	3.44	105	-0.103	0,382
Innovation	52	3.62	53	3.41	105	-0.215**	0,037
BFI-2-S professor							
Extroversion	52	3.79	53	3.85	105	0.050	0,668
Amiability	52	4.22	53	4.02	105	-0.208**	0,033
Responsibility	52	4.12	53	3.82	105	-0.314*	0,059
Negative Emotionality	52	2.53	53	2.44	105	-0.103	0,504
Open-mindedness	52	4.02	53	3.91	105	-0.112	0,441
Autonomous motivation to teach							
External Motivation	52	2.55	53	2.38	105	-0.175	0,307
Unconscious Motivation	52	3.79	53	3.62	105	-0.159	0,308
Motivation Identified	52	4.80	53	4.66	105	-0.142*	0,087
Intrinsic motivation	52	4.61	53	4.64	105	0.018	0,719
BESSI-45 student (Bessi-20 in endline)							
Self-management	988	3.47	982	3.41	1,970	-0.053*	0,056

	Control Group		Treatment Group		Total		
	N	Mean	N	Mean	N	Dif. Of means	p-value
Social Commitment	988	3.25	982	3.20	1,970	-0.043	0,162
Cooperation	988	3.60	982	3.62	1,970	0.019	0,459
Emotional Management	988	3.13	982	3.15	1,970	0.016	0,640
Innovation	988	3.30	982	3.25	1,970	-0.046*	0,096
BFI-2-S Student Index (BFI-2-XS on endline)							
Extroversion	986	3.40	981	3.35	1,967	-0.049	0,105
Amiability	986	3.79	980	3.76	1,966	-0.032	0,263
Responsibility	986	3.36	980	3.31	1,966	-0.051*	0,096
Negative Emotionality	986	2.83	980	2.84	1,966	0.002	0,949
Open-mindedness	986	3.45	980	3.42	1,966	-0.024	0,426
RMET Index							
% RMET Score	984	0.67	977	0.67	1,961	-0.003	0,511
EPOCH							
Commitment	977	3.17	972	3.14	1,949	-0.025	0,428
Perseverance	977	3.63	972	3.55	1,949	-0.092**	0,022
Optimism	977	3.32	972	3.24	1,949	-0.081*	0,079
Connectivity	977	4.02	972	3.99	1,949	-0.031	0,535
Happiness	977	3.74	972	3.71	1,949	-0.042	0,447
Parental Relationship Index							
Caregiver Index...	1,073	4.24	1,065	4.25	2,138	0.022	0,674
Caregiver Index...	1,032	4.15	1,022	4.16	2,054	-0.003	0,666

Note: There is no balance contrast results for the Mindset and Growth Mindset Index for teachers. For students, there is also no equilibrium contrast for the Perseverance Index (GRIT) and Growth Mindset (DMI), standardized mean score of the 5 domains of social-emotional skills of Pentabilities with controls, and measures of self-awareness for social-emotional skills since these indicators were only added in endline. The sample includes all randomized students with baseline and endpoint levels. * p<0.10, **p<0.05, ***p<0.01

The results of the equilibrium contrast in southern Spain, presented in **Table 5**, also do not show significant differences between the different groups.

Table 5: Equilibrium tests between experimental groups (Southern Region)

	Control Group		Treatment Group		Total		
	N	Mean	N	Mean	N	Dif. Of means	p-value
Student Characteristics							
Male	395	0,51	402	0,49	797	-0,014	0,699
Age in 2022	783	13,78	820	13,81	1603	0,021	0,750

	Control Group		Treatment Group		Total		
	N	Mean	N	Mean	N	Dif. Of means	p-value
Spain	756	0,97	783	0,96	1.539	-0,010	0,318
Home Features							
People in the house	787	5,66	818	4,74	1.605	-0,917*	0,060
Bedrooms in the house	787	3,56	818	3,50	1.605	-0,055	0,852
Country of Birth of Caregivers							
Spain	724	0,85	715	0,82	1.439	-0,028	0,151
School Province							
Huelva	219	0,24	199	0,21	418	-0,035	0,391
Malaga	173	0,19	182	0,19	355	-0,002	0,960
Jaén	146	0,16	160	0,17	306	0,005	0,909
Almeria	95	0,10	131	0,14	226	0,031	0,484
Ceuta	97	0,11	111	0,11	208	0,008	0,853
Cadiz	71	0,08	62	0,06	133	-0,014	0,756
Seville	41	0,05	68	0,07	109	0,025	0,595
Granada	42	0,05	37	0,04	79	-0,008	0,862
Melilla	26	0,03	18	0,02	44	-0,010	0,834

Note: The sample includes all randomized students with baseline and endpoint levels. * P<0.10, **P<0.05, ***P<0.01

Table 5 (continued):

Equilibrium tests between experimental groups (Southern Region)

	Control Group		Treatment Group		Total		
	N	Mean	N	Mean	N	Dif. Of means	p-value
Use of Teachers' Time							
Direct instruction to the whole class (Master classes)	38	2.00	42	2.40	80	0.405	0.118
Working with students individually	40	1.75	42	2.12	82	0.369*	0.069
Lead the class in small groups	40	1.75	42	2.17	82	0.417*	0.055
Addressing student disciplinary issues	40	1.60	42	1.62	82	0.019	0.915
Facilitate group work	40	2.02	42	2.19	82	0.165	0.455

	Control Group		Treatment Group		Total		
	N	Mean	N	Mean	N	Dif. Of means	p-value
Addressing student disciplinary issues	40	1.77	42	1.69	82	-0.085	0.665
Meetings with the rest of the teaching staff/management	40	1.75	42	2.00	82	0.250	0.102
Delivery of evaluations	40	1.62	40	1.43	80	-0.200	0.243
Use assessment results	40	1.70	42	1.55	82	-0.152	0.349
Supervisory features	40	1.80	41	1.71	81	-0.093	0.624
Professional development activities	40	2.15	41	2.44	81	0.289	0.240
Preparation for state assessments, etc.	40	1.85	42	1.71	82	-0.136	0.563
Individual planning time	40	2.95	43	3.19	83	0.236	0.333
Communicate with the students' parental figures	40	2.77	42	2.48	82	-0.299	0.219
Collaborative planning time	40	2.67	42	2.64	82	-0.032	0.896
Administrative duties	40	3.50	42	3.26	82	-0.238	0.313
Pre-Instruction	40	2.98	43	3.26	83	0.281	0.265
Post-instruction	40	3.02	42	3.45	82	0.427	0.111
Student's Perception of the Teacher							
It makes me feel like I matter	823	3.85	886	3.80	1,709	-0.047	0.364
Show interest in my work	823	4.17	889	4.14	1,712	-0.038	0.398
Motivates me to succeed	819	4.07	890	3.98	1,709	-0.089*	0.095
It encourages me to explain my process...	820	3.97	887	3.90	1,707	-0.073	0.154
Involve the whole class	823	4.32	890	4.21	1,713	-0.107**	0.024
Praises us for good ideas	823	4.25	886	4.21	1,709	-0.034	0.474
Encourage us to ask/comment...	823	4.20	888	4.22	1,711	0.019	0.688
He criticizes us when we make mistakes	822	2.02	887	1.96	1,709	-0.065	0.290
Encourages student self-management	821	3.93	881	4.04	1,702	0.111**	0.021
Encourages us to help each other	823	4.28	887	4.23	1,710	-0.051	0.287
Autonomous motivation to teach							
External Motivation	41	2.30	43	2.30	84	-0.008	0.959

	Control Group		Treatment Group		Total		
	N	Mean	N	Mean	N	Dif. Of means	p-value
Unconscious Motivation	41	3.33	43	3.35	84	0.020	0.898
Motivation Identified	41	4.70	43	4.67	84	-0.021	0.812
Intrinsic motivation	41	4.43	43	4.40	84	-0.031	0.812

Note: The sample includes all randomized students with baseline and endpoint levels. * p<0.10, **p< 0.05, ***p<0.01

Table 5 (continued):

Equilibrium tests between experimental groups (Southern Region)

	Control Group		Treatment Group		Total		
	N	Mean	N	Mean	N	Dif. Of means	p-value
BESSI-45 professors							
Self-management	41	3.71	42	3.63	83	-0.075	0.533
Social Commitment	41	3.35	42	3.32	83	-0.035	0.778
Cooperation	41	3.93	42	3.95	83	0.026	0.821
Emotional Management	41	3.50	42	3.30	83	-0.200	0.175
Innovation	41	3.29	42	3.48	83	0.186	0.193
BFI-2-S Professor							
Extroversion	41	3.76	43	3.70	84	-0.058	0.672
Amiability	41	4.38	43	4.41	84	0.029	0.807
Responsibility	41	4.03	43	3.94	84	-0.087	0.574
Negative Emotionality	41	2.53	43	2.62	84	0.092	0.561
Open-mindedness	41	3.57	43	3.84	84	0.276	0.106
Autonomous motivation to teach							
External Motivation	41	2.30	43	2.30	84	-0.008	0.959
Unconscious Motivation	41	3.33	43	3.35	84	0.020	0.898
Motivation Identified	41	4.70	43	4.67	84	-0.021	0.812
Intrinsic motivation	41	4.43	43	4.40	84	-0.031	0.812
BESSI 45 student (Bessi-20 in endline)							
Self-management	790	3.37	826	3.43	1,616	0.064**	0.043
Social Commitment	790	3.15	826	3.14	1,616	-0.008	0.811
Cooperation	790	3.63	826	3.63	1,616	0.007	0.815

	Control Group		Treatment Group		Total		
	N	Mean	N	Mean	N	Dif. Of means	p-value
Emotional Management	790	2.92	826	2.99	1,616	0.062	0.103
Innovation	790	3.14	826	3.16	1,616	0.026	0.419
BFI-2-S Student Index (BFI-2-XS on endline)							
Extroversion	789	3.36	825	3.38	1,614	0.019	0.537
Amiability	789	3.84	825	3.84	1,614	0.006	0.850
Responsibility	789	3.49	824	3.52	1,613	0.029	0.396
Negative Emotionality	789	2.89	824	2.86	1,613	-0.024	0.514
Open-mindedness	789	3.47	823	3.47	1,612	-0.005	0.880
RMET Index							
% RMET Score	790	0.66	825	0.66	1,615	-0.001	0.822
EPOCH							
Commitment	782	3.23	815	3.30	1,597	0.070*	0.099
Perseverance	783	3.64	816	3.68	1,599	0.046	0.299
Optimism	782	3.29	815	3.28	1,597	-0.004	0.931
Connectivity	782	4.11	815	4.10	1,597	-0.018	0.674
Happiness	782	3.74	815	3.69	1,597	-0.045	0.340
Caregiver Index							
Caregiver Index 1	847	4.26	865	4.24	1,712	-0.013	0.543
Caregiver Index 2	829	4.13	841	4.16	1,670	0.029	0.209

Note: There is no balance contrast results for the Mindset and Growth Mindset Index for teachers. For students, there is also no equilibrium contrast for the Persistence Index (GRIT) and Growth Mindset (DMI), standardized mean score of the 5 domains of social-emotional skills of Pentabilities with controls, and measures of self-awareness for social-emotional skills since these indicators were only added in endline. The TROS indicator balances, for the sake of brevity, have not been included for the southern region either. The sample includes all randomized students with baseline and endpoint levels. * p<0.10, **p< 0.05, ***p<0.01

4.3 Degree of participation and attrition by groups

This section examines two aspects: the rate of participation in the program and the completeness of the final survey by the participants. Both factors are critical as they directly influence the estimation of program outcomes.

Degree of participation

"Participation" is defined as the proportion of teachers assigned to treatment who have attended initial training and are therefore considered treated. "Compliance" is defined as the proportion of teachers assigned to the treatment who received at least one mentoring session after the initial training. Using these two concepts is critical to understanding the ongoing nature of the treatment

being evaluated. As previously described, treatment consists not only of initial training but also of follow-up through mentoring and focus groups.

The definition of compliance and participation at the student level is linked to the status of the teachers to whom they are assigned.

Starting with the analysis conducted in Catalonia, **Table 6** shows the participation and compliance statistics for this region. Of the total number of teachers assigned to treatment, 97% completed the initial training. The second row of **Table 6** shows statistics on the level of compliance of teachers. 92% of the teachers assigned to the treatment continued with the treatment after the initial training, participating in at least one mentoring session, while only 8% did not conduct any mentoring sessions.

Table 6: Participation record (Catalonia)

	Teacher Level		Student Level	
	%	N	%	N
Participation	96.84 %	92	100.00 %	2,142
Compliance	91.58 %	87	98.60 %	2,112
Total	**100.00 %**	**95**	**100.00 %**	**2,142**

Note: The participation record has been calculated on the total population of teachers and students belonging to the courses that were assigned to the treatment.

Regarding the methodology applied for the Southern section, the definition of these concepts is the same as that provided for Catalonia. The data for this section are shown in **Table 7**, which indicates the degree of participation of teachers in the South region, reflecting that 84% of teachers assigned to the treatment group participated in the intervention.

Table 7: Participation Record (South)

	Teacher Level		Student Level	
	%	N	%	N
Yes participation	83.33 %	60	73.33 %	1,280
Non-participation	16.67 %	12	26.67 %	466
Total	**100.00 %**	**72**	**100.00 %**	**1,747**

Note: The participation record has been calculated on the total population of teachers belonging to the courses that were assigned to the treatment.

Attrition by groups

Starting with Catalonia, **Table 8** shows statistics on the total number of students and teachers who have signed the informed consent, participated in the study, and for whom information has been

obtained through the different surveys. Due to logistical limitations, it is not possible to survey all groups of the study population. In this case, the term population indicates the total number of students who have signed the informed consent and participated in the project, while the potential sample is the subset of this population that has been surveyed. The respondent variable indicates the total number of students and teachers from whom data has been collected through surveys. These statistics have been separated by treatment and control groups, also showing the total.

Table 8: Population with informed consent and final sample (Catalonia)

	Surveyed			Potential Sample			Population		
	TG	CG	Total	TG	CG	Total	TG	CG	Total
Student Level									
Baseline	1,104	1,077	2,181	1,244	1,207	2,451	1,401	1,504	2,905
Midline	1,101	1,082	2,183	1,244	1,207	2,451	1,401	1,504	2,905
Endline	1,082	1,083	2,165	1,244	1,207	2,451	1,401	1,504	2,905
Final activity	1,053	1,064	2,117	1,244	1,207	2,451	1,401	1,504	2,905
Teacher Level									
Baseline	58	63	121	62	67	129	95	89	184
Midline	61	64	125	62	67	129	95	89	184
Endline	59	58	117	62	67	129	95	89	184
Final activity	62	63	125	62	67	129	95	89	184

Note: The population described includes all students and faculty with consent to participate in the study. The potential sample is the group of teachers and students who, for logistical reasons, we seek to survey. The respondents are the group of teachers and students that we finally managed to survey. Baseline: Baseline survey. Midline: Midline survey. Endline: Endline survey.

Regarding attrition, it should first be noted that 'attrition' is defined as the proportion of the total number of subjects (students or teachers) who participate in the project and who abandon it after its start for a clear and manifest reason. The reason may be due to the teacher losing the intention to continue with the treatment, or the school deciding to abandon the project, etc. The attrition percentage is calculated for both the total population participating in the study and for the subset of the sample that is surveyed (**Table 9**). It is worth mentioning that, for both students and teachers, the attrition percentages are low and do not exceed 5% for teachers or 2% for students, whether calculated on the total population or the sample.

Table 9: Attrition and dropout record (Catalonia)

	Treatment		Control		Total	
	%	N	%	N	%	N
Teacher level (Total population)						
Yes	4.21 %	4	4.49 %	4	4.35 %	8
No	95.79 %	91	95.51 %	85	95.65 %	176
Total	**100 %**	**95**	**100 %**	**89**	**100 %**	**184**
Teacher Level (Total Sample)						
Yes	1.61 %	1	4.48 %	3	3.10 %	4

	Treatment		Control		Total	
	%	N	%	N	%	N
No	98.39 %	61	95.52 %	64	96.90 %	125
Total	**100 %**	**62**	**100 %**	**67**	**100 %**	**129**
Level of the student (Total population)						
Yes	1.93 %	27	0.47 %	7	1.17 %	34
No	98.07 %	1,374	99.53 %	1,497	98.83 %	2,871
Total	**100 %**	**1,401**	**100 %**	**1,504**	**100 %**	**2,905**
Student Level (Total Sample)						
Yes	2.17 %	27	0.58 %	7	1.39 %	34
No	97.83 %	1,217	99.42 %	1,200	98.61 %	2,417
Total	**100 %**	**1,244**	**100 %**	**1,207**	**100 %**	**2,451**

Note: The population is considered as the total number of teachers and students participating in the project. The sample is considered as the total number of teachers and students who participate in the project and are selected to be surveyed.

Table 10 and **Table 11** show the analysis for the intervention in southern Spain. As can be seen, the level of attrition in the treatment of teachers and students in southern Spain is close to 8.5%.

Table 10: Population with informed consent and final sample (South)

	Treatment			Control			Total		
	Population	Sample		Population	Sample		Population	Sample	
	N	N	%	N	N	%	N	N	%
Student Level									
Baseline	1,088	982	90.26%	1,007	884	87.79%	2,095	1,866	89.07%
Midline	1,088	885	81.34%	1,007	863	85.70%	2,095	1,748	83.44%
Teacher Level									
Baseline	71	52	73.24%	69	50	72.46%	140	102	72.86%
Endline	71	58	81.69%	69	56	81.16%	140	114	81.43 %
Endline Observer	71	63	88.73%	69	58	84.06%	140	121	86.43%

Note: The population described includes all students and faculty with consent to participate in the study.

Table 11: Attrition Record (South)

	Teacher Level		Student Level	
	%	N	%	N
Yes	8.51 %	12	8.50 %	178
No	91.49 %	129	91.50 %	1,917
Total	**100 %**	**141**	**100 %**	**2,095**

Note: The table describes the level of attrition in the treatment of teachers and students in Sur.

5 Results of the evaluation

Random assignment of the experimental sample to the control and treatment groups ensures that, a sufficiently large sample given, the groups are statistically comparable. Therefore, any differences observed after the intervention can be causally associated with the treatment. Econometric analysis provides, in essence, this comparison. Nevertheless, this analysis has the advantages of allowing other variables to be included to increase accuracy in the estimates and provide confidence intervals for the estimates. In this section, the econometric analysis and the estimated regressions are presented, as well as the analysis of the results obtained.

5.1 Description of econometric analysis: estimated regressions

The main empirical specification estimates the effects of treatment on the intention to treat (ITT), reflecting the causal impact of offering to implement the intervention at an ESO grade on the outcomes of interest. The main outcomes are assessed at the student level, while treatment allocation is randomized at the ESO grade level. Because of this, principal regressions are estimated at the student level, and standard errors are grouped at the grade level to reflect the level of randomization (Abadie et al., 2022). The average effects of the program's treatment on the outcomes of interest are estimated by conditioning on the reference covariates and the fixed effects of the randomization strata in the following general equation:

$$Y_{igst} = \alpha + \boldsymbol{\beta} T_{igs} + \gamma Y_{igs0} + \sum_{w=1}^{W} \delta_i^w + \sum_{c=1}^{4} \eta_s^c + \epsilon_{igst}$$

Where Y_{igst} is the result of interest in the endline, α is a constant, β is the parameter of interest, and Y_{igs0} is the result in the baseline (whenever available, otherwise the term is omitted). Controls are conducted by fixed effects of the week of the year of measurement and by strata of the complexity index and (δ's, η's respectively).

5.2 Analysis of the results

5.2.1 Primary and secondary outcomes

This section presents the results of the assessment on the main and secondary indicators. As in the previous sections, the analysis is approached by differentiating between regions: Catalonia and the South.

The effects on the different blocks described above are developed below.

Teaching actions more focused on promoting the integral development of students

Catalonia

Starting the analysis in Catalonia, the first thing to note is that, after the face-to-face training by the mentors, it is expected that all the treated teachers will implement the proposed pedagogy. Therefore, it is first checked whether the program modifies the teachers' practices and classroom environment as intended. This check can be done based on the secondary results shown in the tables covering teachers' time use inside and outside the classroom for different activities.

The analysis in **Table 12** reveals a notable change in the allocation of teachers' time within the classroom. Surprisingly, treated teachers spend 35 minutes less each week addressing student disciplinary issues compared to their counterparts in the control group. This translates to a total decrease of 42% in the time spent on disciplinary actions each week, a significant difference at the 5% level.

Beyond disciplinary concerns, the treated group also sees a significant reduction in time spent leading small group discussions (1.5 hours) and working with students individually (39 minutes). This shift aligns with pedagogical goals that focus on how the teacher interacts with students and emphasize moving away from traditional teacher-centered approaches toward fostering higher-order thinking, autonomy, and student engagement. The reduction in teacher-led direct discussions and individual attention is consistent with the program's strategy to empower students and foster a learning environment that promotes mutual peer support. These findings, together with the other negative but not significant coefficients, suggest a broad transformation in the teaching practices of the teachers treated towards a more student-centered approach.

Table 13 reveals that the treatment has significant implications beyond the classroom. Treated teachers also experience a reduction in the time spent addressing students' disciplinary problems outside of class, approximately 27 minutes each week (the result of multiplying the regression coefficient, 0.446, by the 60 minutes that make up an hour). This consistent decrease in disciplinary management confirms that the program has a positive impact on reducing disciplinary problems both inside and outside the classroom.

The analysis in **Table 14** explores the time spent on the most recent day on a set of teaching or non-teaching activities. Although none of the coefficients are significant, once baseline levels are controlled, the treated teachers become more involved in daily planning, communication with various stakeholders, and administrative tasks. Finally, when considering the time before and after instruction in **Table 15**, there are again positive coefficients, but not significant. Overall, the results suggest that the intervention is not only efficient in terms of time to implement but also frees up teacher time, reducing the occurrence and management of disciplinary problems without significantly increasing the time spent on alternative tasks.

Continuing with the exploration of the impact on teachers' classroom practices, we then investigate how these changes are perceived from the students' perspective. To measure this, specific

questions are incorporated into student surveys about their lead teacher (enrolled in the program). Initially, the survey includes 11 questions, but for the endline, an additional 11 questions are added aimed at capturing the expected results of a successful adoption of the proposed pedagogy. These additional items focus on capturing whether the teacher contributes to a harmonious and positive classroom environment and whether classroom activities encourage students to reflect on their learning and identify areas for improvement.

This analysis is conducted based on the tables that present the results of the students' perception of the teacher. The analysis is structured to first examine the answers to the initial set of 11 questions and then separately analyze the additional 11 questions added in the endline. The final specification combines all 22 questions.

Table 16 shows the results of the factor analysis of the survey items. Statistically significant and positive effects are found in the three specifications: the treated students rate their teachers more positively, increasing their grades by 0.13-0.14 standard deviations. Interestingly, Table 17 shows that the specification of 11 additional items added to the final survey has a greater effect (0.126), suggesting that teaching strategies and the environment have changed due to the intervention. Overall, the positive coefficients in all models suggest that the intervention is effective in improving students' perceptions of their teachers. This is an important outcome, as teachers' positive perceptions can influence student engagement, motivation, and ultimately their educational success.

The findings of the TROS classroom observation survey, completed by the external observers, are also presented through tables. In general, the results show no noticeable effects, with one exception: the positive and significant coefficient of 'No interaction' in **Table 19** that assesses the nature of the interaction between teachers and students. However, without adjusting for baseline levels, this result is meaningless, suggesting initial imbalances that prevent further interpretations.

Since teachers are regularly trained and mentored in a pedagogy that involves socio-emotional skills, it is also necessary to investigate the impact of the program on their own skills and motivation. In general, no significant trends are observed, apart from two specific subdomains. Treated teachers have lower scores in the Open-mindedness domain, as indicated by the results of the BFI-2-S teacher indicator (**Table 24**). Additionally, the analysis of motivation surveys for autonomous teaching in **Table 25** reveals that the treated teachers exhibit reduced levels of intrinsic motivation. These findings suggest that, although the program's intensive focus on social-emotional pedagogy may influence certain aspects of the teachers' self-perception and motivation, the effects are limited and do not indicate a generalized or systematic impact on their motivation.

Table 12: Use of time (most recent week) in classes by the teacher with controls (Catalonia)

	Lectures	Individually	Driving small groups	Disciplinary	Group work
	(1)	(2)	(3)	(4)	(5)
Treatment	-0.354	-0.653*	-1.562***	-0.583**	-0.279
	(0.38)	(0.38)	(0.59)	(0.22)	(0.56)
Mean	2.34	2.56	3.65	1.39	3.31
Standard deviation	2.26	2.42	3.36	1.51	2.82
N	103	102	102	102	103
Adj. R^2	0.34	0.11	0.00	0.11	0.16

Note: Baseline time use scores are monitored. Fixed effects of strata are included for the level of complexity and fixed effects for the week of the final survey. Standard errors are grouped at the school-grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 13: Use of time (most recent week) spent on activities that are not directly related to classroom instruction by the teacher with controls (Catalonia)

	Disciplinary	Meetings	Evaluations	Results	Supervision	Professional	Prep.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	-0.446*	-0.457	0.077	-0.315	-0.428	-0.589	0.331
	(0.22)	(0.46)	(0.38)	(0.26)	(0.33)	(0.48)	(0.45)
Mean	1.74	3.04	1.63	1.26	1.80	2.41	1.32
Standard deviation	1.61	2.12	1.90	1.56	2.27	2.26	2.16
N	101	102	99	101	103	102	99
Adj. R^2	0.33	0.09	0.08	0.18	0.10	0.17	0.11

Note: Baseline time use scores are monitored. Fixed effects of strata are included for the level of complexity and fixed effects for the week of the final survey. Standard errors are grouped at the school-grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 14: Use of time (the most recent day) spent on activities that are not directly related to classroom instruction by the teacher with controls (Catalonia)

	Planning	Communication	Collaborative Plan.	Administrative	Other
	(1)	(2)	(3)	(4)	(5)
Treatment	0.583	0.108	0.228	0.268	-0.353
	(0.38)	(0.26)	(0.23)	(0.34)	(0.71)
Mean	2.45	1.46	1.61	1.55	1.67
Standard deviation	2.18	1.21	1.85	2.08	3.20
N	103	102	101	101	18
Adj. R^2	0.18	0.26	0.51	0.32	0.75

Note: Baseline time use scores are monitored. Fixed effects of strata are included for the level of complexity and fixed effects for the week of the final survey. Standard errors are grouped at the school-grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 15: Use of time (the most recent day) dedicated to activities by the teacher with controls (Catalonia)

	Pre-Instruction (1)	Post-instruction (2)
Treatment	0.016	0.263
	(0.39)	(0.36)
Mean	2.11	2.36
Standard deviation	1.81	2.20
N	103	101
Adj. R^2	-0.04	0.24

Note: Baseline time use scores are monitored. Fixed effects of strata are included for the level of complexity and fixed effects for the week of the final survey. Standard errors are grouped at the school-grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 16: The scoring factor of students' perception of teachers, with controls (Catalonia)

	11-item (1)	11 extra items (2)	22-item (3)
Treatment	0.126**	0.144*	0.139**
	(0.06)	(0.08)	(0.07)
Mean	-0.09	-0.09	-0.09
Standard deviation	1.04	1.07	1.05
N	1881	1875	1867
Adj. R^2	0.29	0.18	0.30

Note: Baseline survey scores are monitored. Fixed effects of strata are included for the level of complexity and fixed effects for the week of the final survey. Standard errors are grouped at the school-grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 17: The composite score of students' perception of teachers, with controls (Catalonia)

	11-item (1)	11 extra items (2)	22-item (3)
Treatment	0.098**	0.126**	0.111**
	(0.05)	(0.06)	(0.06)
Mean	3.67	3.66	3.66
Standard deviation	0.76	0.86	0.78
N	1915	1915	1915

	11-item	11 extra items	22-item
	(1)	(2)	(3)
Adj. R^2	0.27	0.24	0.28

Note: Baseline survey scores are monitored. Fixed effects of strata are included for the level of complexity and fixed effects for the week of the final survey. Standard errors are grouped at the school-grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 18: TROS: Classroom environment, with controls (Catalonia)

	Class Instruction	Small group instruction	Individual	Other
	(1)	(2)	(3)	(4)
Treatment	0.088	-0.063	-0.050	0.021
	(0.06)	(0.05)	(0.05)	(0.03)
Mean	0.45	0.26	0.22	0.06
Standard deviation	0.30	0.24	0.20	0.14
N	110	110	110	110
Adj. R^2	0.03	0.01	-0.00	-0.01

Note: Baseline survey scores are monitored. Fixed effects of strata are included for the level of complexity and fixed effects for the week of the final survey. Standard errors are grouped at the school-grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 19: TROS: Interaction, with controls (Catalonia)

	No interaction	Instructional	Managerial	Social	Collaborative	Other
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.066**	-0.057	0.034	0.012	-0.030	-0.010
	(0.03)	(0.05)	(0.03)	(0.01)	(0.02)	(0.02)
Mean	0.15	0.59	0.15	0.03	0.05	0.02
Standard deviation	0.16	0.28	0.16	0.07	0.11	0.09
N	110	110	110	110	110	110
Adj. R^2	0.08	0.08	-0.02	-0.10	-0.01	-0.00

Note: Baseline survey scores are monitored. Fixed effects of strata are included for the level of complexity and fixed effects for the week of the final survey. Standard errors are grouped at the school-grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 20: TROS: The nature of the interaction, with controls (Catalonia)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.015	0.044	-0.088	-0.025	0.005	-0.005	0.029	0.029
	(0.02)	(0.03)	(0.05)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)
Mean	0.06	0.12	0.48	0.05	0.07	0.07	0.06	0.10
Standard deviation	0.10	0.16	0.26	0.12	0.12	0.12	0.12	0.16

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
N	110	110	110	110	110	110	110	110
Adj. R^2	-0.01	0.05	0.02	0.04	-0.00	0.00	0.13	-0.05

Note: Baseline survey scores are monitored. Fixed effects of strata are included for the level of complexity and fixed effects for the week of the final survey. Standard errors are grouped at the school-grade level. The elements of the nature of the interaction are as follows: (1) Questioning (process); (2) Questioning (content); (3) Explanation; (4) Commenting (e.g., general discussion of sports); (5) Listening; (6) Give signs or directions; (7) Modeling/demonstration; (8) Other.* p<0.10, **p<0.05, ***p<0.01.

Table 21: TROS: Purpose of interaction I, with controls (Catalonia)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.033	0.067	-0.064	-0.011	-0.023	0.021	0.015
	(0.06)	(0.04)	(0.04)	(0.02)	(0.03)	(0.03)	(0.02)
Mean	0.40	0.28	0.19	0.06	0.09	0.16	0.05
Standard deviation	0.30	0.23	0.20	0.13	0.14	0.18	0.10
N	110	110	110	110	110	110	110
Adj. R^2	0.01	0.01	0.04	-0.04	-0.00	0.04	0.01

Note: Baseline survey scores are monitored. Fixed effects of strata are included for the level of complexity and fixed effects for the week of the final survey. Standard errors are grouped at the school-grade level. The elements of the purpose of the interaction are as follows: (1) Content focus (e.g., subject area content); (2) Process Focus; (3) Focus on the product (e.g., outcome); (4) Present multiple perspectives on the topic; (5) Redirect the student's thinking; (6) Show interest in the student's work; (7) Show personal appreciation for the student. * p<0.10, **p<0.05, ***p<0.01.

Table 22: TROS: Purpose of interaction II, with controls (Catalonia)

	(1)	(2)	(3)	(4)	(5)
Treatment	0.026	-0.028	0.009	0.020	-0.035
	(0.04)	(0.02)	(0.03)	(0.02)	(0.03)
Mean	0.11	0.05	0.09	0.04	0.09
Standard deviation	0.17	0.12	0.16	0.09	0.15
N	110	110	110	110	110
Adj. R^2	0.02	-0.00	0.02	-0.06	0.10

Note: Baseline survey scores are monitored. Fixed effects of strata are included for the level of complexity and fixed effects for the week of the final survey. Standard errors are grouped at the school-grade level. The elements of the purpose of the interaction are as follows: (1) Content focus (e.g., subject area content); (2) Process Focus; (3) Focus on the product (e.g., outcome); (4) Present multiple perspectives on the topic; (5) Redirect the student's thinking; (6) Show interest in the student's work; (7) Show personal appreciation for the student. * p<0.10, **p<0.05, ***p<0.01.

Table 23: BESSI-45 teacher, with controls (Catalonia)

	Self-management	Social Commitment	Cooperation	Emotional Management	Innovation
Treatment	0.065	0.124	-0.049	0.030	0.100
	(0.07)	(0.08)	(0.06)	(0.08)	(0.07)
Mean	3.77	3.41	4.04	3.69	3.63

	Self-management	Social Commitment	Cooperation	Emotional Management	Innovation
Standard deviation	0.48	0.56	0.41	0.59	0.46
N	103	103	103	103	103
Adj. R^2	0.46	0.47	0.45	0.49	0.49

Note: BESSI-45 baseline scores are monitored. Fixed effects of the level of complexity and week of the final survey are included. Standard errors are grouped at the school grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 24: BFI-2-S teacher, with controls (Catalonia)

	Extroversion	Amiability	Responsibility	Neg. Emotion	Mental Apert.
Treatment	-0.017	-0.084	-0.119	-0.001	-0.171**
	(0.08)	(0.07)	(0.09)	(0.09)	(0.08)
Mean	3.77	4.15	4.14	2.51	3.97
Standard deviation	0.62	0.45	0.71	0.61	0.65
N	103	103	103	103	103
Adj. R^2	0.55	0.56	0.67	0.55	0.64

Note: BFI-2-S baseline scores are checked. Fixed effects of the level of complexity and week of the final survey are included. Standard errors are grouped at the school grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 25: Autonomous motivation for teaching, with controls (Catalonia)

	External	Unconscious	Identified	Intrinsic
Treatment	-0.119	-0.101	-0.019	-0.169***
	(0.14)	(0.15)	(0.07)	(0.06)
Mean	2.65	3.73	4.76	4.75
Standard deviation	0.91	0.79	0.32	0.27
N	104	104	104	104
Adj. R^2	0.22	0.19	0.26	0.24

Note: The baseline scores of the Autonomous Motivation are controlled. Fixed effects of the level of complexity and week of the final survey are included. Standard errors are grouped at the school-grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 26: Growth Mindset and Mindset for Teachers, with Controls (Catalonia)

	Growth Mindset Index (1)	Teacher Mindset Index (2)
Treatment	-0.141	0.005
	(0.22)	(0.06)
Mean	2.47	4.36

	Growth Mindset Index	Teacher Mindset Index
	(1)	(2)
Standard deviation	0.96	0.30
N	111	113
Adj. R^2	-0.03	-0.03

Note: Fixed effects of the level of complexity and week of the final survey are included. Standard errors are grouped at the school-grade level. * p<0.10, **p<0.05, ***p<0.01.

Southern Spain

The following tables present the results of the intervention in southern Spain. The analysis does not reveal consistent and statistically significant effects of the treatment on classroom dynamics. Regarding the effect of the intervention in southern Spain on teachers, Table 33 indicates a decrease in the measures reported in four BESSI domains. Additionally, they report a lower level in the Responsibility domain of the BFI-2-S teacher indicator, as presented in Table 34. These findings highlight the important role of treatment duration and the provision of sustained mentoring support, which are essential for teachers to implement the program effectively.

Table 27: Use of time (most recent week) in classes by the teacher with controls (Southern Region)

	Lectures	Individually	Driving small groups	Disciplinary	Group work
	(1)	(2)	(3)	(4)	(5)
Treatment	-0.370	0.386	0.377	0.070	-0.300
	(0.91)	(0.79)	(0.47)	(0.39)	(0.32)
Mean	3.04	2.17	1.68	1.36	2.24
Standard deviation	2.89	2.02	2.23	1.67	2.49
N	80	82	82	82	82
Adj. R^2	0.04	0.03	0.07	0.20	0.18

Note: Baseline time use scores are monitored. Fixed effects of strata are included for Ceuta and Melilla and fixed effects for the week of the final survey. Standard errors are grouped at the school-classroom level. * p<0.10, **p<0.05, ***p<0.01.

Table 28: Use of time (most recent week) spent on activities that are not directly related to classroom instruction by the teacher with controls (Southern Region)

	Disciplinary	Meetings	Evaluations	Results	Supervision	Professional	Prep.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.450	0.080	-0.051	0.151	0.120	-0.278	0.866**
	(0.40)	(0.17)	(0.24)	(0.14)	(0.29)	(0.29)	(0.38)
Mean	1.41	1.62	1.02	0.93	1.40	2.16	0.92

	Disciplinary (1)	Meetings (2)	Evaluations (3)	Results (4)	Supervision (5)	Professional (6)	Prep. (7)
Standard deviation	1.78	1.72	1.37	0.94	1.96	2.28	1.74
N	82	82	80	82	81	81	82
Adj. R^2	0.19	0.14	0.02	0.12	0.06	0.00	0.32

Note: (7) Preparation for the required state, regional and local assessments. Baseline time use scores are controlled. Fixed effects of strata are included for Ceuta and Melilla and fixed effects for the week of the final survey. Standard errors are grouped at the school-classroom level. * p<0.10, **p<0.05, ***p<0.01.

Table 29: Use of time (most recent day) spent on activities that are not directly related to classroom instruction by the teacher with controls (Southern Region)

	Planning (1)	Communication (2)	Collaborative Plan (3)	Administrative (4)
Treatment	-0.178	-0.433	-0.222	-0.149
	(0.48)	(0.40)	(0.27)	(0.33)
Mean	1.83	1.45	1.00	1.84
Standard deviation	1.85	1.75	1.00	1.90
N	82	82	82	82
Adj. R^2	0.00	0.12	0.22	-0.04

Note: Baseline time use scores are monitored. Fixed effects of strata are included for Ceuta and Melilla and fixed effects for the week of the final survey. Standard errors are grouped at the school-classroom level. * p<0.10, **p<0.05, ***p<0.01.

Table 30: Use of time (the most recent day) dedicated to activities by the teacher with controls (Southern Region)

	Pre-Instruction (1)	Post-instruction (2)
Treatment	-0.344	0.250
	(0.26)	(0.31)
Mean	1.70	1.61
Standard deviation	1.64	1.68
N	82	82
Adj. R^2	0.17	0.06

Note: Baseline time use scores are monitored. Fixed effects of strata are included for Ceuta and Melilla and fixed effects for the week of the final survey. Standard errors are grouped at the school-classroom level. * p<0.10, **p<0.05, ***p<0.01.

Table 31: The scoring factor of students' perception of teachers, with controls (Southern Region)

	11-item (1)	11 extra items (2)	22-item (3)
Treatment	0.041	-0.019	0.025
	(0.10)	(0.11)	(0.11)
Mean	0.02	0.03	0.02
Standard deviation	1.01	0.97	1.00
N	1,543	1,522	1,515
Adj. R^2	0.27	0.17	0.21

Note: Baseline survey scores are monitored. Fixed effects of strata are included for Ceuta and Melilla and fixed effects for the week of the final survey. Standard errors are grouped at the school-classroom level. * p<0.10, **p<0.05, ***p<0.01.

Table 32: The composite score of students' perception of teachers, with controls (Southern Region)

	11-item (1)	11 extra items (2)	22-item (3)
Treatment	0.036	-0.000	0.016
	(0.07)	(0.09)	(0.08)
Mean	3.85	3.88	3.86
Standard deviation	0.73	0.82	0.74
N	1,578	1,577	1,578
Adj. R^2	0.24	0.22	0.25

Note: Baseline survey scores are monitored. Fixed effects of strata are included for Ceuta and Melilla and fixed effects for the week of the final survey. Standard errors are grouped at the school-classroom level. * p<0.10, **p<0.05, ***p<0.01.

Table 33: BESSI-45 teacher, with controls (Southern Region)

	Self-management	Social Commitment	Cooperation	Emotional Management	Innovation
Treatment	-0.251***	-0.213***	-0.208***	-0.212***	-0.074
	(0.09)	(0.07)	(0.07)	(0.07)	(0.08)
Mean	3.95	3.63	4.17	3.79	3.59
Standard deviation	0.53	0.65	0.51	0.57	0.61
N	83	83	83	83	83
Adj. R^2	0.50	0.62	0.52	0.68	0.61

Note: BESSI-45 baseline scores are monitored. Fixed effects are included for Ceuta and Melilla and for the week of the final survey. Standard errors are grouped at the school-classroom level. * p<0.10, **p<0.05, ***p<0.01.

Table 34: BFI-2-S teacher, with controls (Southern Region)

	Extroversion	Amiability	Responsibility	Neg. Emotion	Mental Apert.
Treatment	-0.137	-0.010	-0.252***	0.103	0.088
	(0.12)	(0.07)	(0.08)	(0.10)	(0.10)
Mean	3.79	4.37	4.15	2.48	3.68
Standard deviation	0.69	0.50	0.63	0.61	0.68
N	82	82	82	82	82
Adj. R^2	0.52	0.61	0.72	0.56	0.64

Note: BFI-2-S baseline scores are checked. Fixed effects are included for Ceuta and Melilla and for the week of the final survey. Standard errors are grouped at the school-classroom level. * p<0.10, **p<0.05, ***p<0.01.

Table 35: Autonomous motivation to teach, with controls (Southern Region)

	External	Unconscious	Identified	Intrinsic
Treatment	-0.106	-0.072	-0.076*	-0.160
	(0.13)	(0.12)	(0.04)	(0.10)
Mean	2.34	3.34	4.76	4.63
Standard deviation	0.96	0.95	0.28	0.47
N	84	84	84	84
Adj. R^2	0.39	0.43	0.18	0.31

Note: The baseline scores of the Autonomous Motivation are controlled. Fixed effects are included for Ceuta and Melilla and for the week of the final survey. Standard errors are grouped at the school-classroom level. * p<0.10, **p<0.05, ***p<0.01.

Table 36: Growth Mindset and Mindset for Teachers, with Controls (Southern Region)

	DMI Index	Mindset Self-Evaluation: Index
	(1)	(2)
Treatment	-0.110	-0.054
	(0.18)	(0.04)
Mean	2.58	4.38
Standard deviation	1.12	0.29
N	108	108
Adj. R^2	0.03	-0.03

Note: Fixed effects are included for Ceuta and Melilla and for the week of the final survey. Standard errors are grouped at the school-classroom level. * p<0.10, **p<0.05, ***p<0.01.

Improved social-emotional skills

Catalonia

After examining the impacts at the classroom level, the focus shifts to the individual level, particularly on the effects of the program on improving students' awareness and socio-emotional skills in the short term. These were assessed primarily by a standardized activity designed to capture behaviors indicative of social-emotional skills. To measure awareness, novel measures are introduced to understand how individuals perceive their own socio-emotional skills and those of their peers. Finally, to measure changes in socio-emotional skills, external observers' ratings of students' behaviors during the standardized activity are evaluated. All these assessments and measurements are conducted in the form of indicators, as explained in **Section 3.4**.

With respect to behavioral changes within the Pentabilities domains, Table 37 details the external observer ratings derived from the standardized activity. The absence of discernible trends from which to draw meaningful conclusions raises the possibility that the designed standardized activity was not adequate to effectively simulate a classroom environment to elicit behaviors. In an authentic classroom environment, students interact with each other under the supervision of their teachers, providing a rich context to observe behaviors indicative of social-emotional skills. This approach, involving students in a structured activity and employing trained, unbiased external observers to assess these behaviors, may not have adequately captured the subtleties of a real classroom, thus limiting the ability to make informed interpretations about students' actual social-emotional competencies.

Self-reported and teacher-collected data on students' social-emotional skills at the baseline, midline, and endpoint are also examined. **Table 38** and **Table 39** reveal modest improvements for both self-evaluations and teacher observations, respectively. For the former, there is a slight positive effect (significant at the 1% level) in the Innovation domain (**Table 38**). For the latter, the positive and significant coefficients in **Table 39** indicate that, on average, treated students receive higher scores from their teachers in all socio-emotional domains at the end of the period compared to control students, after accounting for their midline scores. Importantly, no faculty observations are collected through BESSI at the baseline, and by the time of the midline surveys, the treated faculty have completed training and are receiving monthly mentoring support.

Therefore, these results should be interpreted with caution, recognizing the change in the socio-emotional skills of the treated students as an improvement from midline to endline.

Finally, although no significant differences are observed at the domain level, a detailed examination of individual behaviors reveals that treated students receive higher scores in a particular aspect of emotional management: recognizing when their approaches are unsuccessful. As the first column of **Table 44** shows, the scores for the treated students are around 3.10, compared to 2.90 for those in the control group.

In summary, although BESSI's findings hint at modest changes at the student level, it is important to recognize that improvements in social-emotional skills—a side effect of changes in teacher behavior to changes in the classroom environment—may not immediately manifest at the individual level, especially over the course of a few months. Such changes often take time to take hold and may not be easily detectable through short-term surveys.

Table 37: Standardized mean score of the 5 domains of socio-emotional skills of Pentabilities with controls (Catalonia)

	Autonomy	Cooperation	Emotional management	Responsibility	Thought
	(1)	(2)	(3)	(4)	(5)
Treatment	0.015	0.042	0.054	0.027	-0.057
	(0.05)	(0.06)	(0.06)	(0.07)	(0.07)
Mean	0.01	0.01	0.00	0.01	0.02
Standard deviation	1.00	1.00	1.00	0.99	1.00
N	1,818	1,821	1,806	1,808	1,806
Adj. R^2	0.00	0.01	0.01	0.01	0.04

Note: Fixed effects of strata are included for the complexity level and fixed effects for the week of the final survey. Standard errors are grouped at the school-grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 38: Self-reported BESSI-20, with controls (Catalonia)

	Self-management	Social Commitment	Cooperation	Emotional Management	Innovation
Treatment	0.012	0.031	-0.010	-0.007	0.095***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Mean	3.54	3.21	3.66	3.22	3.32
Standard deviation	0.69	0.76	0.68	0.80	0.70
N	1,948	1,948	1,948	1,948	1,948
Adj. R^2	0.37	0.42	0.31	0.44	0.36

Note: BESSI-45 baseline scores are monitored. Fixed effects of the level of complexity and week of the final survey are included. Standard errors are grouped at the school grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 39: BESSI-20 observed by teachers, with controls (Catalonia)

	Self-management	Social Commitment	Cooperation	Emotional Management	Innovation
Treatment	0.121***	0.078**	0.110**	0.088**	0.089**
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)

	Self-management	Social Commitment	Cooperation	Emotional Management	Innovation
Mean	3.27	3.15	3.37	3.32	3.28
Standard deviation	1.06	0.95	0.84	0.88	0.86
N	885	885	885	885	885
Adj. R^2	0.75	0.73	0.64	0.67	0.71

Note: BESSI-45 baseline scores are monitored. Fixed effects of the level of complexity and week of the final survey are included. Standard errors are grouped at the school grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 40: BFI-2-XS, with controls (Catalonia)

	Extroversion	Amiability	Responsibility	Neg. Emotion	Mental Apert.
Treatment	0.003	-0.000	-0.012	0.011	0.000
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Mean	3.41	3.68	3.32	3.02	3.40
Standard deviation	0.77	0.69	0.69	0.74	0.75
N	1,928	1,927	1,927	1,927	1,926
Adj. R^2	0.36	0.27	0.24	0.23	0.17

Note: BESSI-45 baseline scores are monitored. Fixed effects of the level of complexity and week of the final survey are included. Standard errors are grouped at the school grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 41: RMET, GRIT & DMI, with controls (Catalonia)

	RMET	Grit	DMI
Treatment	-0.147	0.025	0.073
	(0.21)	(0.04)	(0.07)
Mean	18.31	3.33	3.71
Standard deviation	3.97	0.55	1.15
N	1,933	1,127	1,113
Adj. R^2	0.27	0.00	0.01

Note: Baseline survey scores are controlled for RMET only (Grit and DMI were not included in the initial survey). Fixed effects of the level of complexity and week of the final survey are included. Standard errors are grouped at the school grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 42: Standardized mean score of individual behaviors in the Pentabilities subdomains: Autonomy and initiative, with controls (Catalonia)

	PB B11	PB B12	PB B13	PB B14
Treatment	-0.005	0.039	0.007	-0.011

	PB B11	PB B12	PB B13	PB B14
	(0.05)	(0.05)	(0.05)	(0.05)
Mean	2.90	2.86	2.48	3.03
Standard deviation	1.01	1.08	1.05	0.97
N	1,816	1,810	1,782	1,799
Adj. R^2	0.00	0.00	0.01	0.00

Note: Fixed effects of the level of complexity and the week of the final survey are included. Standard errors are grouped at the school grade level. *p<0.10, **p<0.05, ***p<0.01

Table 43: Standardized mean score of individual behaviors in the Pentabilities subdomains: Cooperation, with controls (Catalonia)

	PB B21	PB B22	PB B23	PB B24	PB B25
Treatment	0.025	0.027	0.034	0.021	0.022
	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)
Mean	2.28	2.88	2.80	3.27	2.84
Standard deviation	0.89	0.91	0.90	0.78	0.81
N	1,818	1,820	1,783	1,809	1,796
Adj. R^2	0.01	0.01	0.01	0.01	0.01

Note: Fixed effects of the level of complexity and week of the final survey are included. Standard errors are grouped at the grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 44: Standardized mean score of individual behaviors in the subdomains of Pentabilities: Emotional management, with controls (Catalonia)

	PB B31	PB B32	PB B33	PB B34
Treatment	0.168**	0.095	0.007	0.038
	(0.06)	(0.06)	(0.04)	(0.06)
Mean	2.91	2.92	3.01	3.03
Standard deviation	0.96	0.94	0.77	0.90
N	1,111	1,198	1,792	1,727
Adj. R^2	0.01	0.02	0.00	0.01

Note: Fixed effects of the level of complexity and the week of the final survey are included. Standard errors are grouped at the school grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 45: Standardized mean score of individual behaviors in the subdomains of Pentabilities: Responsibility, with controls (Catalonia)

	PB B41	PB B42	PB B43	PB B44
Treatment	0.043	0.025	0.013	0.027
	(0.06)	(0.05)	(0.06)	(0.06)
Mean	3.26	3.73	3.24	3.74
Standard deviation	1.01	0.81	0.98	0.80
N	1,803	1,803	1,780	1,773
Adj. R^2	0.01	0.01	0.02	0.02

Note: Fixed effects of the level of complexity and the week of the final survey are included. Standard errors are grouped at the school grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 46: Standardized mean score of individual behaviors in the subdomains of Pentabilities: Thinking skills, with controls (Catalonia)

	PB B51	PB B52	PB B53
Treatment	-0.099*	-0.033	-0.010
	(0.06)	(0.06)	(0.06)
Mean	2.28	2.19	2.52
Standard deviation	0.95	0.88	0.96
N	1,794	1,800	1,803
Adj. R^2	0.03	0.03	0.02

Note: Fixed effects of the level of complexity and week of the final survey are included. Standard errors are grouped at the school grade level. * p<0.10, **p<0.05, ***p<0.01.

Southern Spain

The results of the impact of the treatment on students in southern Spain are presented below. The analysis does not reveal consistent and statistically significant effects of the treatment on student outcomes. Although there are a few isolated significant coefficients, such as the level of perseverance of the students in **Table 50**, these findings do not form a coherent pattern from which definitive conclusions can be drawn.

Table 47: Self-reported BESSI-20, with controls (Southern Region)

	Self-management	Social Commitment	Cooperation	Emotional Management	Innovation
Treatment	0.023	0.028	-0.014	0.046	0.024
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Mean	3.47	3.16	3.72	3.04	3.25

	Self-management	Social Commitment	Cooperation	Emotional Management	Innovation
Standard deviation	0.69	0.73	0.65	0.82	0.72
N	1,616	1,616	1,616	1,616	1,616
Adj. R^2	0.42	0.41	0.32	0.46	0.42

Note: BESSI-45 baseline scores are monitored. Fixed effects of the level of complexity and week of the final survey are included. Standard errors are grouped at the school grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 48: BESSI-20 observed by teachers, with controls (Southern Region)

	Self-management	Social Commitment	Cooperation	Emotional Management	Innovation
Treatment	0.039	-0.005	-0.062	0.021	0.087
	(0.10)	(0.09)	(0.09)	(0.08)	(0.10)
Mean	3.15	3.15	3.51	3.33	3.18
Standard deviation	1.02	0.95	0.81	0.82	0.89
N	1,042	1,042	1,042	1,042	1,042
Adj. R^2	0.02	0.01	0.03	0.03	0.02

Note: Fixed effects are included for Ceuta and Melilla and for the week of the final survey. Standard errors are grouped at the school-classroom level. * p<0.10, **p<0.05, ***p<0.01.

Table 49: BFI-2-XS student, with controls (Southern Region)

	Extroversion	Amiability	Responsibility	Neg. Emotion	Mental Apert.
Treatment	0.001	0.005	-0.036	0.006	-0.045
	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)
Mean	3.34	3.74	3.42	3.12	3.55
Standard deviation	0.75	0.71	0.72	0.75	0.78
N	1,574	1,577	1,576	1,573	1,571
Adj. R^2	0.32	0.22	0.26	0.22	0.13

Note: BFI-2-S baseline scores are checked. Fixed effects are included for Ceuta and Melilla and for the week of the final survey. Standard errors are grouped at the school-classroom level. * p<0.10, **p<0.05, ***p<0.01.

Table 50: RMET, GRIT & DMI, with controls (Southern Region)

	RMET	Grit	DMI
Treatment	-0.346	0.070*	0.037
	(0.21)	(0.04)	(0.08)
Mean	18.05	3.26	3.70

	RMET	Grit	DMI
Standard deviation	4.33	0.56	1.22
N	1,617	1,687	1,674
Adj. R^2	0.22	0.01	0.00

Note: Baseline survey scores are controlled for RMET only (Grit and DMI were not included in the initial survey). Fixed effects are included for Ceuta and Melilla and for the week of the final poll. Standard errors are grouped at the school-classroom level. * p<0.10, **p<0.05, ***p<0.01.

Improvement in behaviors demonstrated in class in active pedagogy environments

Catalonia

Table 51 presents the results on self-awareness of socio-emotional skills during the standardized activity, revealing that there are no significant changes in individual perceptions of their own socio-emotional competencies. Similarly, no significant effects were found on awareness of peers' socio-emotional skills, as detailed in **Table 52**. The near-zero effect sizes could be due to this novel theoretical framework, which contrasts internal and external viewpoints, being an unproven approach that may not effectively capture the intended results.

Table 51: Measure of self-awareness of the level of socio-emotional skills with controls (Catalonia)

	Autonomy	Cooperation	Emotional management	Responsibility	Thought
	(1)	(2)	(3)	(4)	(5)
Treatment	0.045	0.134	0.018	-0.095	0.034
	(0.11)	(0.13)	(0.13)	(0.15)	(0.10)
Mean	1.26	1.48	1.93	2.21	1.53
Standard deviation	1.80	2.14	2.31	2.63	1.96
N	1,778	1,752	1,770	1,775	1,778
Adj. R^2	0.01	0.01	0.00	0.00	0.00

Note: Fixed effects of strata are included for the complexity level and fixed effects for the week of the final survey. Standard errors are grouped at the school-grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 52: Measure of awareness of the socio-emotional skills of peers during a standardized activity, with controls (Catalonia)

	Autonomy	Cooperation	Emotional management	Responsibility	Thought
Treatment	-0.000	0.061	-0.057	-0.128	-0.057
	(0.09)	(0.14)	(0.14)	(0.13)	(0.10)
Mean	1.26	1.73	2.25	2.20	1.91

	Autonomy	Cooperation	Emotional management	Responsibility	Thought
Standard deviation	1.32	1.78	1.98	2.07	1.69
N	1,742	1,735	1,737	1,741	1,742
Adj. R^2	0.02	0.02	0.01	0.01	0.01

Note: Fixed effects of strata are included for the complexity level and fixed effects for the week of the final survey. Standard errors are grouped at the school-grade level. * p<0.10, **p<0.05, ***p<0.01.

Other results

Tables 53 and 54 show the impact on the indicators of well-being (Commitment, Perseverance, Optimism, Connectivity and Happiness) and on the parental relationship index, respectively. The results presented in the tables do not show a significant effect of the treatment on the observed indicators.

Table 53: EPOCH, with controls (Catalonia)

	Commitment	Perseverance	Optimism	Connectivity	Happiness
Treatment	-0.021	0.051	0.005	0.009	-0.008
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Mean	3.28	3.61	3.42	4.12	3.80
Standard deviation	0.93	0.89	1.01	0.89	0.98
N	1,926	1,926	1,926	1,926	1,925
Adj. R^2	0.25	0.42	0.41	0.28	0.45

Note: EPOCH baseline scores are monitored. Fixed effects of the level of complexity and week of the final survey are included. Standard errors are grouped at the school grade level. * p<0.10, **p<0.05, ***p<0.01.

Table 54: The parental relationship index, with controls (Catalonia)

	Primary caregiver (1)	Secondary caregiver (2)
Treatment	0.012	0.025
	(0.02)	(0.02)
Mean	4.24	4.15
Standard deviation	0.45	0.50
N	1,887	1,740
Adj. R^2	0.12	0.17

Note: Parental Relationship Index baseline scores are monitored. Fixed strata effects are included for the complexity level. Standard errors are grouped at the school grade level. * p<0.10, **p<0.05, ***p<0.01.

Southern Spain

As in the implementation in Catalonia, the results of the project in southern Spain also do not suggest a significant positive impact on the indicators of well-being (**Table 55**) or on the indicators of parental relationship (**Table 56**).

Table 55: EPOCH, with controls (Southern Region)

	Commitment	Perseverance	Optimism	Connectivity	Happiness
Treatment	-0.025	-0.061*	0.009	0.021	0.017
	(0.04)	(0.03)	(0.05)	(0.03)	(0.05)
Mean	3.30	3.73	3.44	4.23	3.80
Standard deviation	0.90	0.91	1.02	0.88	0.96
N	1,586	1,588	1,586	1,586	1,586
Adj. R^2	0.23	0.48	0.41	0.26	0.43

Note: EPOCH baseline scores are monitored. Fixed effects are included for Ceuta and Melilla and for the week of the final survey. Standard errors are grouped at the school-classroom level. * p<0.10, **p<0.05, ***p<0.01.

Table 56: The parental relationship index, with controls (Southern Region)

	Primary caregiver	Secondary caregiver
	(1)	(2)
Treatment	-0.017	0.032*
	(0.02)	(0.02)
Mean	4.25	4.14
Standard deviation	0.45	0.47
N	1,572	1,503
Adj. R^2	0.15	0.13

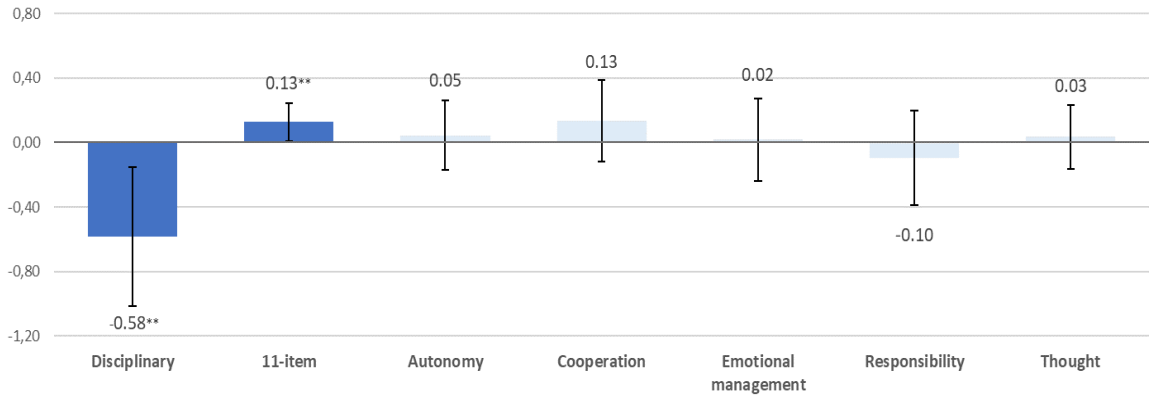
Note: Parental Relationship Index baseline scores are monitored. Fixed effects are included for Ceuta and Melilla and for the week of the final survey. Standard errors are grouped at the school-classroom level. * p<0.10, **p<0.05, ***p<0.01.

6 Conclusions of the evaluation

The findings of this study provide evidence on the impact of integrating formative assessment of students' socio-emotional skills into teachers' classroom practices. Recognizing the critical role of non-cognitive skills in both academic success and future outcomes in the labor market (especially for students from disadvantaged backgrounds), this pedagogy brings these skills to the fore within the natural classroom environment. The program offers teachers practical practice that helps them create a common language around socio-emotional skills. They also learn to integrate them into classroom dynamics and to provide feedback to students based on behavioral evidence collected.

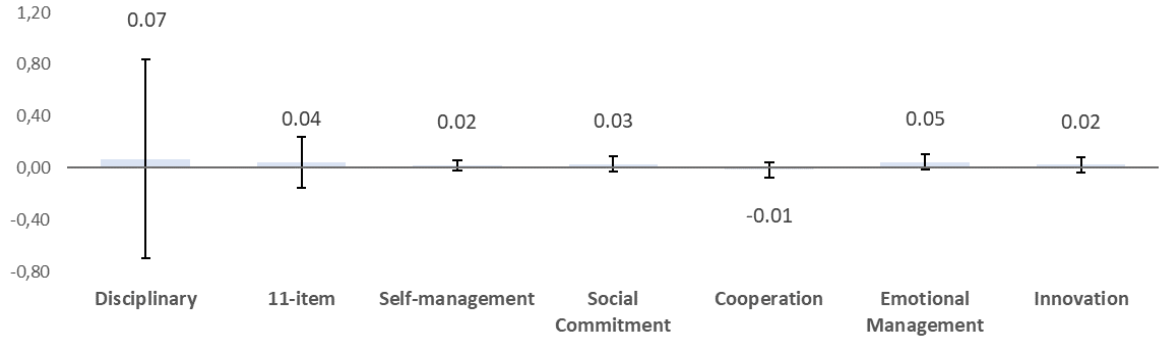
The intervention is implemented through two randomized controlled trials in Catalonia and southern Spain (Andalusia, Ceuta, and Melilla).

Figure 5: Effect of the intervention on the main indicators (Catalonia)



Note: Indicators whose treatment effect is significant at 5% are presented in dark blue, and those indicators that are not significant at 10% are presented in light blue. The effects included in the graphics refer to regressions with controls.

Figure 6: Effect of the intervention on the main indicators (Southern Region)



Note: Indicators whose treatment effect is significant at 5% are presented in dark blue, and those indicators that are not significant at 10% are presented in light blue. The effects included in the graphics refer to regressions with controls.

The study finds that, applied as a pilot program over a 5-month period with 5 mentoring sessions, Pentabilities has had a significant impact on teachers' classroom practices. The intervention affects teachers' time allocation and improves students' perceptions of their teachers. The significant decrease in time spent on disciplinary issues – 42% in and 26% outside the classroom – suggests an improvement in classroom management, potentially allowing teachers to reallocate their time towards more effective teaching practices and foster a positive learning environment. Notably, the faculty of the treatment group do not report significant changes in preparation time, indicating that beyond the initiation phase, the program is not time-intensive to implement. The impact of the program is further evidenced by the improved perceptions that students have of their teachers, indicating a strengthened teacher-student relationship.

In terms of assessing students' socio-emotional skills, the project uses surveys and develops a novel standardized activity to elicit and rate behaviors. Teachers' favorable evaluations of their students' socio-emotional skills highlight the potential effectiveness of the program. However, no significant changes in students' consciousness or socio-emotional skill levels measured through the standardized activity are captured. This result highlights the challenge of measuring overall changes in short-term socio-emotional skills. The prolonged nature of socio-emotional skill development, as well as the limitations of short-term assessments, suggest that capturing the full extent of these changes may require longer observation periods.

From a policy perspective, this study contributes to the debate on how to integrate non-cognitive skills into schools with disadvantaged youth. Likewise, this program helps highlight the role of educational programs in individual development through interventions specifically aimed at these skills. These results provide evidence that modifying teachers' classroom practices can be an effective way to create more positive learning environments that can gradually affect students and change their behaviors. This finding is particularly significant in addressing the educational needs of disadvantaged youth, offering a promising path for future interventions aimed at reducing gaps in education.

Bibliography

Al-Jbouri, E., Andrews, N. C. Z., Peddigrew, E., Fortier, A., & Weaver, T. (2023). Building elementary students' social and emotional skills: A randomized control trial to evaluate a teacher-led intervention. *School mental health*, *15*(1), 138–150. https://doi.org/10.1007/s12310-022-09538-x

Barnes, T. N., Smith, S. W., & Miller, M. D. (2014). School-based cognitive-behavioral interventions in the treatment of aggression in the United States: A meta-analysis. *Aggression and Violent Behavior, 19*(4), 311–321. https://doi.org/10.1016/j.avb.2014.04.013

Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V., Lawson, J., et al. (2001). Are intuitive physics and intuitive psychology independent? a test with children with asperger syndrome. *Journal of developmental and learning disorders*, 5(1):47–78. https://www.researchgate.net/publication/237239741_Are_intuitive_physics_and_intuitive_psychology_independent_A_test_with_children_with_Asperger_Syndrome

Borghans, L., Lee Duckworth, A., Heckman, J., Weel, B. (2008). The Economics and Psychology of Personality Traits. *Journal of Human Resources*, October 2008, 43 (4) 972-1059; DOI: https://doi.org/10.3368/jhr.43.4.972

Clarke, A.M., Morreale, S., Field, C.A., Hussein, Y., Barry, M.M. (2015). What works in enhancing social and emotional skills development during childhood and adolescence? A review of the evidence on the effectiveness of school-based and out-of-school programs in the UK. A report produced by the WHO Collaborating Center for Health Promotion Research, national university of Ireland Galway. https://assets.publishing.service.gov.uk/media/5a809c17e5274a2e87dbaca5/What_works_in_enhancing_social_and_emotional_skills_development_during_childhood_and_adolescence.pdf

Coelho, V. A., Marchante, M., & Sousa, V. (2015). "Positive attitude": A multilevel model analysis of the effectiveness of a social and emotional learning program for Portuguese middle school students. *Journal of Adolescence, 43,* 29–38. https://doi.org/10.1016/j.adolescence.2015.05.009

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/ES/TXT/?uri=legissum:4617909

Corcoran, R.P., Cheung, A, Kim, E., Xie, C. (2018). Effective Universal school-based social and emotional learning programs for improving academic achievement: A systematic review and meta-analysis of 50 years of research. *Educational Research Review*. https://doi.org/10.1016/j.edurev.2017.12.001

Deming, D. (2017). The Growing Importance of Social Skills in the Labor Market. *The Quarterly Journal of Economics*, Volume 132, Issue 4, November 2017, Pages 1593–1640, https://doi.org/10.1093/qje/qjx022

Duckworth, A. L. and Quinn, P. D. (2009). Development and validation of the short grit scale (grit–s). *Journal of personality assessment*, 91(2):166–174. https://doi.org/10.1080/00223890802634290

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. Child Development, 82, 405-432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Dweck, C. S. (2000). Growth mindset. Self-theories: Their role in motivation, personality, and development. Psychology press

European Commission, Directorate-General for Education, Youth, Sport and Culture, Cefai, C., Bartolo, P., Cavioni, V. (2018). *Strengthening social and emotional education as a core curricular area across the EU: a review of the international evidence: analytical report*, Publications Office. https://data.europa.eu/doi/10.2766/664439

Government of Spain, Ministry of Social Rights and Agenda 2030 (2022). State Action Plan for the Implementation of the European Child Guarantee (2022-2030). Approved on July 5, 2022. https://www.mdsocialesa2030.gob.es/derechos-sociales/infancia-y-adolescencia/Garantia_infantil.htm

Greenberg, M. T., Kusche, C. A., Cook, E. T., & Quamma, J. P. (1995). Promoting emotional competence in school-aged children: The effects of the PATHS curriculum. *Development and Psychopathology, 7*(1), 117–136. https://doi.org/10.1017/S0954579400006374

"The Education Hub. (2020). Mindset self-evaluation for teachers. Retrieved from https://theeducationhub.org.nz/wp-content/uploads/2020/03/Mindset-self-evaluation-for-teachers.pdf"

Heckman, J. and García, J.L. (2017). *Quantifying the Life-cycle Benefits of a Prototypical Early Childhood Program*. Department of Economics. The University of Chicago, May 26, 2017. https://heckmanequation.org/wp-content/uploads/2017/12/abc_comprehensivecba_JPE-SUBMISSION_2017-05-26a_sjs_sjs.pdf

Heckman, J. and Rubinstein, Y. (2001). "The Importance of Noncognitive Skills: Lessons from the GED Testing Program." *American Economic Review*, 91(2):145–149. https://www.aeaweb.org/articles?id=10.1257/aer.91.2.145

Kautz, T., Heckmann, J.J., Diris, R., Weel, B., Borghans, L. (2014). *Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success*. OECD Education Working Papers, No. 110, OECD Publishing. https://www.oecd.org/education/ceri/Fostering-and-Measuring-Skills-Improving-Cognitive-and-Non-Cognitive-Skills-to-Promote-Lifetime-Success.pdf

Kern, M. L., Benson, L., Steinberg, E. A., & Steinberg, L. (2016). The EPOCH Measure of Adolescent Well-Being. *Psychological Assessment, 28*(5), 586–597. https://doi.org/10.1037/pas0000201

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research, 86*(3), 643–680. https://doi.org/10.3102/0034654315626799

Organic Law 2/2006, of 3 May, on Education. https://www.boe.es/eli/es/lo/2006/05/03/2/con

Organic Law 3/2020, of 29 December, amending Organic Law 2/2006, of 3 May, on Education. https://www.boe.es/eli/es/lo/2020/12/29/3/con

Mahoney, J.L., Durlak, J.A., Weissberg, R.P. (2018). An update on social and emotional learning outcome research. *Phi Delta Kappan*. 100(4):18–23. https://doi.org/10.1177/0031721718815668

Newman, J., & Dusenbury, L. (2015). Social and emotional learning (SEL): A framework for academic, social, and emotional success. In *Prevention science in school settings: Complex relationships and processes* (pp. 287-306). New York, NY: Springer New York. https://psycnet.apa.org/record/2015-56845-014

Palczyńska, M., Świst, K. (2018). Personality, cognitive skills and life outcomes: evidence from the Polish follow-up study to PIAAC. Large-scale Assess Educ 6, 2. https://doi.org/10.1186/s40536-018-0056-z

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

Roth, G., Assor, A., Kanat-Maymon, Y., and Kaplan, H. (2007). Autonomous motivation for teaching: how self-determined teaching may lead to self-determined learning. *Journal of educational psychology*, 99(4):761. https://doi.org/10.1037/0022-0663.99.4.761

Schonert-Reichl, K. A., Smith, V., Zaidman-Zait, A., & Hertzman, C. (2012). Promoting children's prosocial behaviors in school: Impact of the "Roots of Empathy" program on the social and emotional competence of school-aged children. *School Mental Health: A Multidisciplinary Research and Practice Journal, 4*(1), 1–21. https://doi.org/10.1007/s12310-011-9064-7

Schonert-Reichl, K.A., Oberle, E., Stewart Lawlor, M., Abbott, D., Thomson, K., Oberlander, T.F., Diamond, A. (2015). Simple-to-administer mindfulness-based school program for elementary school children: A randomized controlled trial. *Developmental Psychology*. 2015; 51(1):52–66. https://doi.org/10.1037/a0038454

Sklad, M., Diekstra, R., De Ritter, M., Ben, J., & Gravesteijn, C. (2012). Effectiveness of school-based universal social, emotional, and behavioral programs: Do they enhance students' development in skill, behavior, and adjustment? *Psychology in the Schools, 49*(9), 892–909. https://doi.org/10.1002/pits.21641

Soto, C. J. and John, O. P. (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, 68:69–81. https://doi.org/10.1016/j.jrp.2017.02.004

Soto, C. J., Napolitano, C. M., Sewell, M. N., Yoon, H. J., & Roberts, B. W. (2022). An integrative framework for conceptualizing and assessing social, emotional, and behavioral skills: The BESSI.*Journal of Personality and Social Psychology, 123*(1), 192–222. https://doi.org/10.1037/pspp0000401

Taylor, R., Oberle, E., Durlak, J.A.; Weissberg, R.P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development*, 88(4), 1156-1171. https://pubmed.ncbi.nlm.nih.gov/28685826/

OCDE (2015). *Skills for Social Progress: The Power of Social and Emotional Skills*. Paris, France: OECD. https://www.oecd.org/education/skills-for-social-progress-9789264226159-en.htm

Waxman, H. C., Wang, M. C., Lindvall, C. M., and Anderson, K. A. (1990). *Teacher Roles Observation Schedule technical manual (revised ed.)*. Center for Research in Human Development and Education, Philadelphia, PA: Temple University. https://core.ac.uk/download/pdf/79650908.pd

Wigelsworth, M., Lendrum, A., Oldfield, J., Scott, A., ten Bokkel, I., Tate, K., & Emery, C. (2016). The impact of trial stage, developer involvement and international transferability on universal social and emotional learning program outcomes: A meta-analysis. *Cambridge Journal of Education, 46*(3), 347–376. https://doi.org/10.1080/0305764X.2016.1195791

Appendix

Economic and regulatory management

1. Introduction

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

Investment 7 "Promotion of Inclusive Growth through the linkage of 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 inclusion model based on the Minimum Income Scheme (MIS), aimed at reducing income inequality and poverty rates. To achieve this goal, the development of pilot projects for the implementation of social inclusion itineraries with communities and autonomous communities, local entities, and Third Sector organizations of Social Action, as well as with various social actors, has been proposed.

Royal Decree 938/2021, dated October 26, which regulates the direct granting of subsidies from the Ministry of Inclusion, Social Security, and Migrations in the field of social inclusion, for an amount of 109,787,404 euros, within the framework of the Recovery, Transformation, and Resilience Plan[14], contributed to achieving critical milestone (as stated in the Council's Implementation Decision) number 350 for the first quarter of 2022 "Improving the access rate of the Minimum Income Scheme, and increasing the effectiveness of the MIS through inclusion policies," which, according to its description, will translate into supporting the socio-economic inclusion of MIS beneficiaries through itineraries: eight collaboration agreements signed with subnational public administrations, social partners, and Third Sector organizations of Social Action to execute the itineraries. These partnership agreements aim to i) improve the access rate to the MIS; ii) increase the effectiveness of the MIS through inclusion policies. Likewise, along with Royal Decree 378/2022, of May 17[15] , it contributed to meeting tracking indicator number 351.1 in the first quarter of 2023 "at least 10 additional collaboration agreements signed with subnational public administrations, social partners, and Third Sector organizations of Social Action to

[14] https://www.boe.es/diario_boe/txt.php?id=BOE-A-2021-17464

[15] https://www.boe.es/diario_boe/txt.php?id=BOE-A-2022-8124

implement pilot projects supporting the socio-economic inclusion of MIS beneficiaries through itineraries", linked to the Operational Arrangements document[16].

In addition, after the implementation and evaluation of each of the subsidized pilot projects, an assessment will be conducted to evaluate the coverage, effectiveness, and success of the minimum income schemes. The publication of this evaluation, which will include specific recommendations to improve the access rate to the benefit and enhance the effectiveness of social inclusion policies, contributes to the achievement of milestone 351 of the Recovery, Transformation, and Resilience Plan scheduled for the first quarter of 2024.

In accordance with Article 3 of Royal Decree 938/2021, of October 26, the granting of subsidies will be conducted by means of a resolution accompanied by an agreement from the person holding the position of Minister of Inclusion, Social Security, and Migrations 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 **August 31, 2022**, the Jaume Fundació Bofill was notified of the Resolution of the General Secretariat for Inclusion and Social Welfare Objectives and Policies granting a subsidy of 12,500,000 euros to the Jaume Fundació Bofill and, on **September 1, 2022**, an Agreement was signed between the General State Administration, through the General Secretariat for Inclusion and Social Welfare Objectives and Policies and the Jaume Fundació Bofill for the implementation of a social inclusion project within the framework of the Recovery, Transformation and Resilience Plan, which was published in the "Boletín Oficial del Estado" on **September 16, 2022** (BOE no. 223)[17].

2. Timeline of the intervention

Article 17(1) of Royal Decree 378/2022 dated May 17, established that the deadline for the implementation of the social inclusion itinerary pilot itineraries subject to 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 comply with the milestones set by the Recovery, Transformation and Resilience Plan with regard to social inclusion policies.

Within this general timeframe, the implementation begins on **November 25, 2022,** with the start of the intervention itinerary, continuing the execution tasks until **November 30, 2023**, and then developing only dissemination and evaluation tasks of the project until **March 31, 2024**.

3. Relevant agents

[16] 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

[17] https://www.boe.es/boe/dias/2022/09/16/pdfs/BOE-A-2022-15164.pdf

Among the relevant agents for the implementation of the project can be mentioned:

- **Fundació Jaume Bofill**, as the beneficiary entity and coordinator of the project.
- The **Ministry of Inclusion, Social Security and Migration (MISSM)** as the sponsor of the project, and as the main responsible for the RCT evaluation process. The General Secretariat for Inclusion (SGI) assumes the following commitments:
    a) Assist the beneficiary entity in the design of the activities to be carried out for the implementation and monitoring of the object of the subsidy, as well as for the profiling of the potential participants of the pilot project.
    b) Design the randomized controlled trial (RCT) methodology of the pilot project in coordination with the beneficiary entity.
    c) Evaluate the pilot project in coordination with the beneficiary entity.
- The **José Manuel Lara Foundation** subcontracted for the implementation of the project in Andalusia, Ceuta, and Melilla.
- **Participating educational centers** supported by public funds, as well as **teachers, tutors, students, and their families**.
- The **Institute of Political Economy and Governance (IPEG),** as a relevant institution for the design and execution of the RCT evaluation of the project.
- **CEMFI and J-PAL Europe**, as scientific and academic institutions supporting MISSM in the design and RCT evaluation.