From Industrial Floors to Classroom Doors: Translating Kaizen Principles into Educational Impact

Pedro Costa de Almeida

Master's Dissertation

Supervisor at FEUP: Prof. José António Barros Basto

U.PORTO

FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Master in Industrial Engineering and Management

2025-06-27

Abstract

This dissertation presents an innovative approach to applying continuous improvement methodologies - rooted in the *Kaizen* philosophy - to the public education system in Portugal. In response to persistent challenges related to equity, learning outcomes, and student engagement, it proposes a structured, scalable, and data-driven intervention model capable of sustainably transforming school practices.

Anchored in a multidimensional theory of change, the project developed and tested a performance measurement model across several schools within an urban cluster facing high socio-educational complexity. The model integrates real-time data dashboards, key performance indicators (KPIs), and a participatory monitoring and evaluation framework. Over several years, metrics such as academic performance, absenteeism, and retention rates were tracked using both longitudinal and cross-sectional strategies.

Through a quasi-experimental design - combining simultaneous control groups and historical comparisons - a robust counterfactual analysis of the intervention's effects was carried out. The results demonstrated significant improvements in academic achievement, reduced absenteeism, and increased student motivation. The Social Return on Investment (SROI) analysis confirmed that the social value generated clearly exceeded the implementation costs.

The dissertation highlights the strategic potential of adapting continuous improvement tools from the industrial sector to the educational context, while also identifying limitations related to data access, ethical constraints, and contextual diversity. It argues that sustainable educational innovation requires embedding a culture of monitoring, feedback, and local ownership of improvement processes.

This work contributes not only a replicable model for educational improvement, but also a broader vision of the school as a living system - where each small change, grounded in evidence and purpose, can drive systemic transformation.

Resumo

Esta dissertação apresenta uma abordagem inovadora à aplicação de metodologias de melhoria contínua - inspiradas na filosofia *Kaizen* - ao sistema de ensino público em Portugal. Perante desafios persistentes relacionados com a equidade, os resultados de aprendizagem e o envolvimento dos alunos, propõe-se um modelo de intervenção estruturado, escalável e baseado em dados, capaz de transformar práticas escolares de forma sustentada.

Ancorado numa teoria da mudança multidimensional, o projeto desenvolveu e testou um modelo de medição de desempenho em várias escolas de um agrupamento inserido num contexto urbano de elevada complexidade socioeducativa. O modelo integra dashboards de dados em tempo real, indicadores-chave de desempenho (KPIs) e uma estrutura participativa de monitorização e avaliação. Foram acompanhadas, ao longo de vários anos, métricas como o desempenho académico, o absentismo e as taxas de retenção, utilizando estratégias longitudinais e de comparação transversal.

Através de um desenho quase-experimental - combinando grupos de controlo simultâneos e comparações históricas - foi possível realizar uma análise contrafactual robusta dos efeitos da intervenção. Os resultados demonstraram melhorias significativas no desempenho académico, redução do absentismo e aumento da motivação dos alunos. A análise do Retorno Social do Investimento (SROI) confirmou um valor social gerado claramente superior aos custos de implementação.

A dissertação evidencia o potencial estratégico da adaptação de ferramentas de melhoria contínua do setor industrial à realidade educativa, identificando também limitações relacionadas com o acesso a dados, restrições éticas e diversidade contextual. Argumenta-se que a inovação educativa sustentável exige o enraizamento de uma cultura de monitorização, feedback e apropriação local dos processos de melhoria.

Este trabalho contribui não só com um modelo replicável de melhoria educativa, mas também com uma visão mais ampla da escola como sistema vivo - onde cada pequena mudança, fundamentada em evidência e propósito, pode gerar transformação sistémica.

UN Sustainable Development Goals

The United Nations Sustainable Development Goals (SDGs) provide a global framework to achieve a better and more sustainable future for all. It includes 17 goals to address the world's most pressing challenges, including poverty, inequality, climate change, environmental degradation, peace, and justice.

The *Kaizen Education Movement* demonstrates a clear alignment with several Sustainable Development Goals (SDGs) established in the United Nations 2030 Agenda for Sustainable Development (United Nations, 2015).

The specific Sustainable Development Goals mentioned have the following names:

SDG 4 Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all

SDG 10 Reduce inequality within and among countries

SGD	Target	Contribution	Performance Indicators and Metrics
4	4.1	By implementing evidence-based strategies and continuous monitoring, the initiative directly supports improvements in student achievement and academic performance.	Increase in students' average grades
	4.2	The use of tailored interventions and real-time feedback enables early identification and support for at-risk students, resulting in fewer students failing.	Reduction in the percentage of failing grades among students
	4.3	Enhanced student engagement and improved classroom practices help decrease absenteeism, ensuring more consistent participation in learning.	Reduction in the total number of absences among students
10	10.1	By focusing on TEIP schools and using granular data to identify and assist students facing disadvantage, the initiative helps reduce educational inequalities and break cycles of social exclusion	Sustained increase in student outreach

Acknowledgements

To my family, who support me every day and motivate me to always strive to be better. If I am proud of who I am today, it is thanks to the education they gave me and the example they continue to be.

To my girlfriend, who has supported me and made me a better person for so many years. She taught me never to give up, to shine, and to appreciate the great little things in life.

To my supervisor at *Kaizen Institute*, João Guedes, for always being present (even on weekends) throughout this journey - for showing consistent interest in the work and in me - but above all, for his character and determination in creating such an important and meaningful initiative as Kaizen Education. May this be the beginning of a very fruitful partnership.

To the entire *Kaizen* team, who always guided me on the right path, gave me all the support I needed, and welcomed me with open arms over the past few months.

To my supervisor at FEUP, Professor José Barros Basto, for all his support throughout this journey, and also to Professor Henriqueta Novoa, for her help in solving the challenges that arose and for opening new paths to knowledge and success.

Finally, a word to myself - for having completed yet another stage with distinction and for always striving to give my best at every step. Fortunately, I can say that I am very proud of the final result and that I truly believe it will, in some way, contribute to society.

To all of you, my most sincere thanks.

"The function of education is to teach one to think intensively and to think critically. Intelligence plus character - that is the goal of true education."

Martin Luther King Jr.

Contents

1 Introduction **1**
- 1.1 Project Context . . . 2
  - 1.1.1 Kaizen Education Movement . . . 2
  - 1.1.2 Kaizen Philosophy in Education . . . 4
- 1.2 Methodology . . . 5
- 1.3 Objectives . . . 8
- 1.4 Document Structure . . . 8

2 Literature Review **10**
- 2.1 Foundational Concepts: Impact and Measurement in Education . . . 10
- 2.2 The Current State of Continuous Improvement Projects in Education . . . 12
  - 2.2.1 Defining Continuous Improvement in Education . . . 13
  - 2.2.2 The Evidence Gap in Continuous Improvement: Scope, Duration, and Design Limitations . . . 13
- 2.3 Culture of Evaluation and Capacity Building in Organisations . . . 14
- 2.4 Towards Evidence-Based Educational Innovation . . . 15
- 2.5 The Growing Culture of Impact Measurement in Portuguese Education . . . 16
- 2.6 Impact Measurement Models and Approaches in a National Context . . . 16
  - 2.6.1 Lack of Baselines, Control Groups, and Validated Scales . . . 17
  - 2.6.2 Current Limitations of Impact Measurement in Portugal . . . 17

3 Problem Description **19**
- 3.1 From Business Efficiency to Educational Excellence: The Evolution of the Kaizen Excellence in Education . . . 19
- 3.2 The Importance of Impact Measurement in Social Projects . . . 21
- 3.3 Current educational situation and its economic relevance . . . 23
- 3.4 The true value of Data . . . 24
- 3.5 Alignment with international recommendations . . . 25

4 Solution **27**
- 4.1 Rethinking Impact Measurement: From Institutional Metrics to Student-Level Insights . . . 27
- 4.2 Definition of the Impact Model . . . 28
  - 4.2.1 Implementation Potential and Systemic Influence of the Impact Measurement Model . . . 29
  - 4.2.2 From Pilot to National Framework: Cost-Based Rationale . . . 30
  - 4.2.3 Toward Scalable, Policy-Aligned Implementation . . . 31
  - 4.2.4 The Impact Measurement Approach . . . 31
  - 4.2.5 Selecting the Right KPIs . . . 32

4.3 Kaizen Education Impact Assessment: Applying the Model 34
4.3.1 Methodological Design and Rationale 34
4.3.2 Limitations and Adaptive Solutions 35
4.3.3 Assumptions 35
4.3.4 Filtering Criteria and Scope of Analysis 37
4.3.5 KPIs Calculation Logic 38
4.3.6 Report and Monitoring Charts 39

5 Results **41**
5.1 Aggregated Results in AE Cerco do Porto 41
5.2 SROI 44
5.2.1 Kaizen Education SROI 45

6 Next Steps **49**
6.1 Quantitative vs Qualitative Indicators 49
6.2 The Difference-in-Differences (DiD) Statistical approach 50

7 Conclusion **53**

References **55**

A Auxiliary calculations of the model's KPIs **60**

B Dashboard views **62**

C SROI Calculations **65**

D NPS Results **67**

E Examples of Kaizen Practices in Action **69**

Acronyms and Symbols

AECP - Agrupamento de Escolas do Cerco do Porto
AIR - American Institute of Research
CIE - CounterFactual Impact Evaluation
CI - Continuous Improvement
DID - Difference-in-Differences
EEF - Education Endowment Foundation
EMIS - Education Management Information System
IMF - International Monetary Fund
KBS - Kaizen Business System
KEE - Kaizen Excellence in Education
KPI - Key Performance Indicator
M&E - Monitoring and Evaluation
NGO - Non-Governmental Organisations
OECD - Organisation for Economic Co-operation and Development
OECD-DAC - OECD-Development Assistance Committee
PDCA - Plan, Do, Check, Act
PDSA - Plan, Do, Study, Act
PISA - The Programme for International Student Assessment
SROI – Social Return on Investment
TEIP - Territórios Educativos de Intervenção Prioritária
UNESCO - United Nations Educational, Scientific and Cultural Organization

List of Figures

1.1 *Kaizen Education* organisational structure 2

3.1 *Kaizen Excellence in Education* Roadmap 20

4.1 Flow of data from extraction to presentation 34

5.1 Evolution of Control and Kaizen for Average Grades, % of Failing Grades and Total Absences per student per year 43

B.1 Initial Dashboard Mock-up - Annual View 62
B.2 Initial Dashboard Mock-up - Longitudinal View 63
B.3 Evolution and comparison of Kaizen vs Control from 2021/2022 to 2023/2024 . . 63
B.4 Evolution and comparison of Kaizen vs Control in 2024/2025 (1st semester) . . 64
B.5 Longitudinal overview of Kaizen vs 5 years past 64

D.1 Net Promoter Score (NPS) Results at Escola Básica do Falcão (2020/2021 to 2022/2023) 67
D.2 Net Promoter Score (NPS) Results at Escola Básica do Falcão (2020/2021 to 2022/2023) - Detailed View 68

E.1 Student filling in KPI board 69
E.2 Student filling in KPI board (2) 70
E.3 Daily Kaizen Dashboard 70
E.4 Daily Kaizen Dashboard (2) 71
E.5 Daily Kaizen Dashboard (3) 71

List of Tables

5.1 Kaizen vs Control Improvement 44
5.2 Proxy values used in the Kaizen Education SROI analysis 47
5.3 SROI Benefit per Component 47
5.4 SROI Calculation Summary 47

A.1 Summary of average grades by year and group (Control vs Kaizen) 60
A.2 Summary of failing grades by year and group (Control vs Kaizen) 60
A.3 Summary of absences by year and group (Control vs Kaizen) 61
A.4 Total Kaizen Absences per student per Academic Year 61
A.5 Total Kaizen Absences Reduced (73%) per student per Academic Year 61
A.6 Total Kaizen Students per Academic Year 61
A.7 Total Saving per Academic Year (€) 61

C.1 Average annual salary increase by level of education 65
C.2 Weighted average of salary increase per additional year of schooling 65
C.3 Retention costs and total cost by level of education (2022) 65
C.4 Retention cost in the public sector by educational cycle (2022) 66
C.5 Summary of retention costs and budget ratio 66
C.6 Key Teaching Cost Indicators (2022) 66

Chapter 1

Introduction

The *Kaizen Education Movement* is an initiative designed to implement continuous improvement methodologies in various educational contexts. Its aim is to reduce absenteeism, improve academic success, increase student and teacher engagement, and foster a culture of ongoing learning and process optimization within schools.

The challenge lies in designing an impact measurement model for the *Movement*, capable of providing a continuous, data-driven understanding of its effectiveness in multiple educational settings.

In any continuous improvement initiative, the success of implementation is heavily dependent on two critical elements: the clarity and coherence of the strategy and the robustness of the mechanisms for monitoring impact. This work was designed to ensure that the impact measured in previous years can be consistently assessed in the future, preserving both the structure and the scalability of the project.

In this context, it is imperative to elucidate certain terms employed throughout the dissertation. Research is defined as the process of developing a thesis, that is, academic work undertaken with the objective of exploring, analyzing, and proposing solutions to a problem within the field of education. The term 'movement' is equivalent to the concept of program and is represented here by *Kaizen Education* as a sector of education in which the concept of continuous improvement is applied.

Kaizen Education aims to transform and optimize educational processes in a gradual and sustained manner, with a focus on continuous improvement of results. The ultimate goal of this intervention is to apply *Kaizen* principles to transform the institution's school reality, aligning it with the continuous improvement practices explored in the thesis.

1.1 Project Context

1.1.1 Kaizen Education Movement

Kaizen philosophy, which originated in Japan, is a fundamental concept in quality management and continuous improvement of organizational processes. The term '*Kaizen*' literally means 'change for the better' or 'continuous improvement', and is based on the idea that constant, systematic small improvements can lead to significant progress over time (Imai, 2012). *Kaizen* is more than just a management technique; it represents an organizational culture that is focused on efficiency, employee involvement, and waste elimination at all levels of the company. Therefore, it can be applied not only to industrial production but also to any area where processes can be analyzed and improved, thereby promoting a culture of operational excellence and continuous innovation.

Kaizen Education by Kaizen Institute is an innovative movement that emerged in Portugal in 2021 at the Agrupamento de Escolas do Cerco do Porto as an internal, non-profit initiative. The organisational structure of the *Kaizen Education Movement* has been summarised in the diagram shown (Figure 1.1), which illustrates its hierarchical and functional structure. At the apex of the structure is the movement itself, which encompasses a set of different projects that correspond to each of the school implementations.

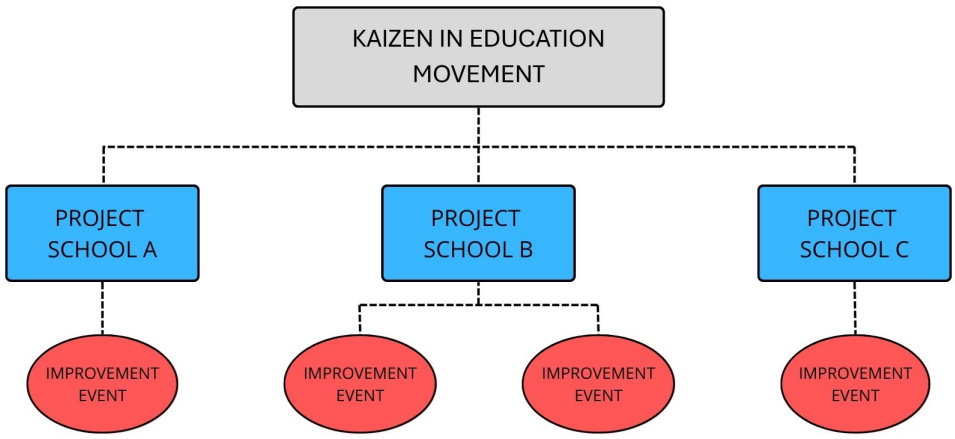


Figure 1.1: *Kaizen Education* organisational structure

Consequently, each school project encompasses numerous enhancement events, comprising concrete, systematised actions intended to optimise pedagogical and organisational processes, adapted to the institution's specific requirements. This model facilitates the integration of a global

perspective with locally and contextually adapted interventions, thereby promoting the dissemination of effective practices and the sustainability of change.

Initially focused on implementing continuous improvement methodologies within a single school context, the movement has since expanded to several other school clusters in the districts of Porto, Lisbon, and Guarda. This expansion is driven by the urgent need to address persistent challenges in the Portuguese education system. Structural issues such as educational inequality, high absence rates, and suboptimal learning conditions remain prevalent - particularly in TEIP (Territórios Educativos de Intervenção Prioritária) schools. These schools serve communities marked by socioeconomic disadvantage and often operate under resource constraints that hinder both teaching and learning.

Several alarming indicators underline the critical nature of these challenges:

- Approximately 30% of students report disliking school, revealing a significant engagement problem (Graham et al., 2022).

- School failure and grade retention are estimated to cost the Portuguese education system over €503 million annually (Calculated with DGEEC (2023); EduStat (2022)).

- It is estimated that the economic value of an additional year of schooling, as a percentage of GDP per capita, can generate an increase of 35% (Psacharopoulos and Patrinos, 2018).

- Only 9% of teachers feel their profession is adequately recognised, reflecting low professional morale (OECD, 2020).

- On average, teachers spend 53% of their time away from direct interaction with students, highlighting inefficiencies in the current organisational structure (OECD, 2020).

In this context, the *Kaizen Education Movement* promotes the application of structured, sustainable improvement methodologies capable of responding to these complex challenges. While rooted in industrial and corporate environments, *Kaizen's* principles - centred on incremental, data-informed change - offer significant potential when adapted to educational contexts. By integrating *Kaizen's* philosophy of incremental and sustainable transformation, this research aims to contribute a strategic, scalable approach to educational improvement - one that is both responsive to local school realities and aligned with national educational goals.

Building on this momentum, the Movement submitted a candidacy to Portugal Inovação Social, aiming to further strengthen its operational capacity and deepen its social impact. This candidacy not only represents a strategic step toward sustainability and growth but also serves as a powerful validation of the movement's relevance and innovation in addressing systemic challenges in education.

1.1.2 Kaizen Philosophy in Education

The *Kaizen Education* methodology is rooted in the principles of continuous improvement, originally developed by Masaaki Imai in 1985 and embedded in the Toyota Management System. The *Kaizen Institute*, with its global presence and over 1,000 consultants, has expanded these principles to various sectors, including education. The core aim of *Kaizen* is to increase productivity, enhance service quality, and improve motivation while minimizing waste and inefficiency. In Portugal, since 1999, *Kaizen Institute* Portugal has helped national organizations implement continuous improvement strategies tailored to the specific needs of their institutions.

In the context of *Kaizen*, *added value* refers to all activities that transform a product or service to fulfil customer needs, for which customers are willing to pay. In other words, it is any step in the process that contributes directly to delivering value to the end consumer. It is essential to identify what adds value to optimise processes and eliminate anything unnecessary.

Conversely, *waste* (or 'muda' in Japanese) is anything that consumes resources - such as time, materials, energy and space - without generating value for the customer. *Kaizen* proposes systematically identifying and eliminating the seven main types of waste: overproduction, waiting time, unnecessary transport, overprocessing, excessive inventory, unnecessary movement and defects (Imai, 2012). By reducing this waste, organisations can become more efficient, competitive and sustainable.

At its core, *Kaizen* emphasizes small, incremental improvements that, over time, lead to significant changes. This methodology fosters a collaborative culture, where all stakeholders - teachers, students, administrators, and staff - work together to achieve efficiency and effectiveness. In schools, *Kaizen* aims to develop structured strategies that improve teaching, enhance student engagement, and optimize administrative processes.

The *methodology* is structured around four main components:

1. Foundations: This is the cultural and strategic groundwork necessary for successful continuous improvement. Key elements include *leadership alignment*, a *shared vision*, and a *commitment* to fostering learning and development at all levels of the school. Strong foundations ensure that the entire school community is engaged in the improvement process.

2. *Kaizen* Change Model: The model guides the change process in schools. It begins with diagnosing areas for improvement, followed by prioritizing issues, designing solutions, and implementing them systematically. Crucially, the process involves continuous monitoring and evaluation to assess progress and make necessary adjustments, creating a cycle of reflection and improvement.

3. Pillars of Improvement: Kaizen Education focus on key areas that are critical to their success. These include:

    - Increasing student engagement through interactive and student-centered teaching practices.

- Analyzing customized key performance indicators to enable data-driven decisions.
- Fostering a culture of daily improvement through small, consistent gains.
- Effective time management to reduce waste and optimize productivity.
- Focusing on value-added tasks that directly impact student learning and well-being.

These pillars guide the strategic actions taken throughout the continuous improvement process.

4. Digitalization: In today's educational environment, technology plays a vital role in *Kaizen's* implementation. By integrating digital tools like dashboards and real-time data tracking systems, schools can monitor progress, track key performance indicators, and make data-driven decisions. This digital approach facilitates continuous feedback and allows schools to quickly adapt strategies to ensure consistent improvement.

By applying these four components, *Kaizen* supports schools in making small but impactful changes that lead to sustainable improvements. With a focus on data-driven decisions and ongoing monitoring, the *Movement* ensures that educational institutions can track progress, address challenges proactively, and ultimately enhance student success. This research aims to show how the structured application of *Kaizen* methods can create a culture of continuous learning and improvement in the education sector.

1.2 Methodology

To initiate this construction process, it became essential to understand what already exists in this field, both in terms of the literature specializing in impact measurement in social projects, as well as through the practical analysis of similar initiatives that have already implemented evaluation methodologies. A comprehensive literature review was therefore conducted about social impact measurement models. In addition, several national and international initiatives that have been recognized for their good practices in this area were analyzed.

Some of the cases analysed include *Teach for Portugal*, *Apps for Good*, *Movimento Transformers*, *Mentes Empreendedoras* and *EPIS - Empresários Pela Inclusão Social*. The study of these examples will focus on the measurement methodologies adopted, the indicators used, the results recorded, and the monitoring and evaluation mechanisms implemented.

The process of creating a social assessment project in this area is comprised of several interconnected stages. Initially, a comprehensive identification and analysis of pertinent national and international references is undertaken. The definition of key performance indicators (KPIs) is guided by the six evaluation criteria set out by OECD (2021): *relevance*, *coherence*, *effectiveness*, *efficiency*, *impact*, and *sustainability*.

The criteria provide a comprehensive framework for the measurement of both the implementation and the outcomes of an intervention. The *relevance* of an initiative is determined by its alignment with actual educational needs, while its *coherence* is evaluated in relation to existing

policies and programmes. The *effectiveness* of the programme is measured through two key indicators: academic improvements and stakeholder engagement. *Efficiency* is assessed by measuring the extent to which resources are used to achieve results. *Impact* focuses on broader changes, such as the influence on educational practices or policy, while *sustainability* considers the programme's long-term viability and potential for scale. Collectively, these dimensions facilitate a rigorous and strategic evaluation process.

The model's structuring process results in the construction of a functional prototype that integrates several essential dimensions. In addition to the creation of a dashboard for the purpose of data visualisation, the design of the main functionalities and the mapping of the relevant data sources, an impact measurement model will be developed that rigorously defines what should be collected, when and how often over the course of each school year.

This model incorporates two distinct components: firstly, continuous monitoring mechanisms that facilitate real-time observation of initiatives' progression; and secondly, impact assessment elements designed to evaluate outcomes over both the medium and long term. A standardised system for the collection and storage of information will also be implemented, ensuring the comparability of data over time and between different projects. This will make it possible to generate robust evidence on the effectiveness of interventions.

In order to ensure the model's cohesiveness, the following issues require resolution or ongoing analysis: the model's long-term sustainability; strategies for ensuring scalability without compromising quality; adjustment of the model to different school and socio-economic contexts; risks associated with collecting and processing sensitive data in an educational environment; and the use of data collected, ensuring it is used ethically, geared towards improving the system, and not as a tool for penalisation. These issues are of fundamental importance and should be given due consideration throughout the project's development to facilitate continuous improvement.

Design of the Movement's Impact Measurement Model

The project involved the development of a comprehensive evaluation framework for the *Kaizen Education Movement* at large.

1. Acting as an External Consultant to the Movement
   The research role was carried out with a degree of externality and critical distance, allowing for an independent analysis of the movement's evolution, decisions, and strategic coherence.

2. Literature Review on Social Impact Evaluation
   The methodological foundation of this research was built upon a comprehensive review of both academic and grey literature on social impact assessment, with a particular focus on methodologies commonly used in education and social innovation. This review provided essential insights into how impact is evaluated across various sectors and helped shape the approach for assessing the effectiveness of the *Kaizen Education movement*.

One of the key frameworks examined in this study is Social Return on Investment (SROI), which provides a structured approach to measuring the social value created by an intervention relative to the resources invested. Commonly applied in social and educational initiatives, the SROI framework allows for the quantification of non-financial outcomes - such as improved wellbeing, educational achievement, and community development - alongside traditional performance indicators. In the context of this project, SROI serves as a valuable tool for assessing the broader impact of continuous improvement strategies in education, enabling a more comprehensive understanding of both academic and social returns (Martinez and Hayes, 2013).

3. Data-Driven Analysis of Educational Performance
   In addition to the SROI model, the literature review underscored the importance of incorporating quantitative indicators to evaluate educational performance and outcomes. For this research, statistical data from various institutional and national sources were analysed to gain insights into the functioning of the education system in Portugal. These data sources helped identify patterns of inefficiency and informed the development of targeted improvement strategies. The use of indicators was essential to understanding the broader challenges faced by schools, particularly in areas such as academic achievement, student engagement, and retention rates.

4. Benchmarking of Educational Social Projects
   The study incorporated an examination of prevailing national initiatives centred on educational and social innovation. Each initiative was examined in terms of its approach to impact measurement, enabling the identification of common standards, effective methodologies, and innovative practices that can inform the development of a comprehensive and adaptable impact measurement model.

5. Consultation with Other Movements
   Semi-structured interviews and collaborative discussions were held with leaders of educational movements, enabling the collection of qualitative data on practical challenges, data usage, ethical concerns, and the role of impact measurement in shaping strategy.

6. Definition of Global KPIs
   A set of Key Performance Indicators (KPIs) was developed to enable systematic assessment of the movement. These were grouped into three dimensions: Academic Impact, Student Engagement and Scalability.

7. Design of Outputs and Communication Tools
   Based on the KPI framework, a series of visual and analytical tools was designed, including dashboards, executive summaries, and visual infographics to facilitate stakeholder engagement and knowledge dissemination.

8. Implementation of the Model
   The process concluded with the creation of a functional prototype for the impact evaluation model. This included mapping relevant data sources, defining data collection routines, and setting up a storage and processing system, while anticipating possible constraints related to data sensitivity, resource availability, and ethical use.

The structure of this methodology ensures that the work is both contextually anchored and theoretically robust. The alignment between local practice and national ambition is central to the *Kaizen Education Movement*'s theory of change and to the overall purpose of this research.

1.3 Objectives

The main objective of this research is to develop and implement a model for continuous improvement in education at the national level, by contributing to the design of an impact measurement framework for the broader *Kaizen Education Movement*. The approach is designed to ensure that the methods and tools applied are context-sensitive, scalable, and sustainable over time.

1. Develop a data collection and analysis system to evaluate the national-level impact of the *Kaizen Education Movement*, providing an integrated view of its influence on school performance, student success, and institutional development.

2. Design dashboards and dissemination tools to support monitoring and knowledge sharing across schools, making data accessible and actionable for practitioners, decision-makers, and funding entities.

In addition, this research seeks to contribute to the broader debate on the improvement of teaching and school governance by demonstrating how structured continuous improvement methodologies can generate measurable gains in educational outcomes and institutional effectiveness.

1.4 Document Structure

This dissertation is structured into three main parts:

- **Part I** – Context and Foundations (Chapters 1–2)
  Introduces the *Kaizen Education Movement*, its relevance to the Portuguese education system, and outlines the research objectives and methodology. It also provides a literature review on continuous improvement, impact measurement, and evaluation culture in education.

- **Part II** – Model Development and Implementation (Chapters 3–5)
  Describes the problem context, details the design of the impact assessment model, and presents its application in a real school setting. Results from the pilot implementation -

including performance indicators and a Social Return on Investment (SROI) analysis - are discussed.

- **Part III** – Future Outlook and Conclusions (Chapters 6–7)
  Offers recommendations for scaling, refining the model, and advancing research. The conclusion summarizes key insights and contributions to policy and practice.

Additional technical details are included in the appendices:

- Appendix A: KPI Calculations
- Appendix B: Dashboard Views
- Appendix C: SROI Data
- Appendix D: NPS Results
- Appendix E: Examples of Kaizen Practices in Action

Chapter 2

Literature Review

This chapter is organised in a sequential and logical way, starting with the key concepts and the conceptual framework, with special emphasis on the importance of impact and its measurement in the educational context (2.1). From this basis, the current state of continuous improvement projects in education is explored (2.2), leading to the need for a more precise definition of this concept (2.2.1) and the identification of its empirical limitations (2.2.2). This is followed by an analysis of the role of evaluation culture and organisational capacity building as facilitating factors for improvement (2.3), culminating in the transition to an evidence-based educational innovation perspective (2.4). In the last sub-chapters, the focus shifts to the Portuguese context, addressing the growth of the impact assessment culture (2.5) and the models used at national level (2.6), ending with a critical reflection on the structural limitations still present, such as the lack of baselines, control groups and validated instruments (2.6.1 and 2.6.2). This structure allows for a coherent progression from theory to practice and from the global to the specific, laying the foundations for subsequent research.

In any rigorous research endeavour, particularly within the complex and value-laden domain of education, the precise articulation of foundational concepts is not merely a formal requirement; it is a conceptual necessity. The following section therefore seeks to ground the study in a robust conceptual foundation by exploring five interrelated constructs: *impact*, *education*, *measurement*, *impact measure*, and *impact measurement in education*. In doing so, it also distinguishes between the universal conception of education and its contextual manifestation in Portugal, offering both theoretical precision and contextual relevance to the subsequent analysis.

2.1 Foundational Concepts: Impact and Measurement in Education

Conceptual ambiguity, particularly the uncritical borrowing of terms from policy language or related fields, can lead to mismatched research methods and misleading conclusions. Therefore, before moving on to the empirical and analytical aspects of impact evaluation, it is essential to establish a common language and clear definitions of the key concepts involved. This is especially

important in educational contexts, where a wide range of stakeholders - policy-makers, educators, researchers, and students - may use the same terms with slightly different meanings or intentions.

The concept of *Impact* refers to the net effect of a specific intervention, policy, or program that can be causally attributed to its implementation, rather than to external or coincidental factors (Rossi et al., 2004). It occupies a central place in program evaluation and differs fundamentally from outputs or outcomes by requiring a demonstrable causal link between the intervention and observed changes. Impact may be positive or negative, direct or indirect, intended or unintended, and may operate at individual, institutional, or systemic levels (Scriven, 1991). Establishing such impact necessitates the use of rigorous methodological controls to distinguish the true effects of the intervention from background noise or confounding influences.

At the heart of educational inquiry lies the concept of *Education* itself, which can be understood as a socially organized and institutionally supported process through which individuals acquire and develop knowledge, cognitive skills, emotional competencies, and values essential to personal fulfilment and active citizenship (UNESCO, 2015). Education encompasses not only formal schooling but also lifelong learning and informal experiences that contribute to human capital and social cohesion. While this is a universal definition, the practical manifestations of education vary significantly across national systems.

In Portugal, education is both a constitutional right and a legal obligation, with compulsory schooling from age six to eighteen (Ministério da Educação, 2023). The system is organized into basic and secondary levels, followed by tertiary education. Over recent decades, Portugal has made substantial progress in improving access and attainment; however, challenges remain in aligning its educational outcomes with OECD benchmarks. For example, only 55% of adults aged 25 to 64 have completed upper-secondary education, compared to the OECD average of 79% (OECD, 2024).

Moreover, educational investment per student remains below the OECD average, with expenditure at USD 11,752 per student compared to the average of USD 14,209 (OECD, 2024). Despite these constraints, Portugal distinguishes itself through a strong policy focus on equity and inclusion, promoting universal access, curricular flexibility, and the integration of students with special educational needs into mainstream education (OECD, 2022b). These priorities reflect both national aspirations and alignment with broader European educational goals.

Measurement, as applied in educational research, refers to the systematic process of assigning numerical or categorical values to abstract constructs according to established rules, with the purpose of enabling structured analysis and informed decision-making as said by Stevens (1946). For such measurement to be meaningful, it must satisfy conditions of validity - that is, accurately capturing the construct of interest - and reliability, ensuring consistent results under similar conditions (Cronbach, 1971).

In the educational domain, Measurement can focus on cognitive domains like literacy and numeracy, as well as affective and behavioral indicators such as motivation, resilience, and attendance. One of the key challenges lies in capturing these multidimensional and dynamic traits

through standardized instruments while remaining sensitive to the cultural and contextual realities of learners, as Brookhart (2013) defined.

Within this framework, an *Impact Measure* is understood as a specific indicator or set of indicators used to evaluate the effects of a program or intervention relative to a baseline or comparator group. These indicators can take quantitative forms, such as test scores, dropout rates, or employment statistics, or qualitative ones, including student and teacher narratives or community perceptions (Rossi et al., 2004). The value of an impact measure hinges on its alignment with the theory of change behind the intervention, the availability and quality of data, and the robustness of the evaluation design (Pawson, 2003). In empirical studies, impact measures often serve as dependent variables in causal models, requiring not only statistical competence but also theoretical grounding and contextual interpretation to assess whether observed changes are meaningful and sustainable.

In *Education* specifically, *Impact Measurement* refers to the systematic evaluation of how policies, interventions, or pedagogical approaches influence learning outcomes and educational equity over time. Historically, such measurements have prioritized standardized academic assessments. However, there is a growing recognition that this narrow focus does not fully capture the broader goals of education, which include socio-emotional development, civic engagement, digital literacy, and overall student well-being (Vincent-Lancrin et al., 2019; Fullan, 2016). Effective impact measurement in educational contexts requires careful attention to both design and timing. In the early phases of program implementation, quasi-experimental methods, such as matched control groups, offer valuable insights into causal relationships, as maintained by Shadish et al. (2002). As programs scale up, and pure control groups become less feasible, alternative strategies such as historical comparisons and longitudinal tracking become essential for capturing systemic trends and long-term effects.

2.2 The Current State of Continuous Improvement Projects in Education

Over the last decade, *Continuous Improvement* (CI) has emerged as a compelling framework for educational reform, rooted in the principles of data-driven inquiry and practitioner-led experimentation. Unlike traditional top-down reforms that emphasize fidelity to externally designed, evidence-based interventions, CI encourages educators to identify localized problems of practice, test solutions in context, and iteratively refine their approaches based on real-time data (Garet et al., 2021). Drawing inspiration from industrial quality improvement models - most notably Deming's Plan-Do-Study-Act (PDSA) cycles - and recent educational adaptations like Bryk et al. (2015) improvement science, CI aims to cultivate a culture of disciplined inquiry and adaptive learning in schools.

While the conceptual foundations of CI are well-developed and widely endorsed (Bryk et al., 2015; Park et al., 2013; Langley et al., 2009), empirical research on its application in educational settings is relatively sparse. A major review by Garet et al. (2021), commissioned by the American

Institutes for Research, sought to synthesize the state of CI implementation by examining 34 empirical studies conducted between 1995 and 2020. Their analysis reveals a landscape marked by enthusiasm and promise, but also by significant limitations in scale, rigor, and impact. This review draws on their findings to critically assess the current state of CI projects in education, exploring their definitions, implementation patterns, outcomes, and underlying challenges.

2.2.1 Defining Continuous Improvement in Education

The literature reveals substantial variability in how Continuous Improvement (CI) is conceptualized and applied in schools. While many studies adopt frameworks grounded in the PDSA cycle Deming (1986); Langley et al. (2009), others use the term more generically to describe any ongoing improvement effort. Garet et al. (2021) found that only 27 of the 34 studies they reviewed explicitly defined CI in a way consistent with iterative inquiry processes. When clearly articulated, definitions often referenced either Deming's cycle or Bryk et al. (2015) six principles of improvement science, which include problem-specific inquiry, system analysis, measurement for improvement, disciplined experimentation, and learning through networks.

The boundaries between CI and related practices - such as data use, professional learning communities, and implementation science - are frequently blurred. This conceptual ambiguity can hinder both implementation fidelity and the accumulation of generalizable knowledge. To qualify for inclusion in the American Institue of Research (AIR) review, studies had to report on inquiry cycles conducted by educators, thereby filtering out looser interpretations of CI that lacked systematic testing of interventions.

2.2.2 The Evidence Gap in Continuous Improvement: Scope, Duration, and Design Limitations

Most CI studies to date have been relatively modest in scope and duration. Of the 34 studies reviewed by Garet et al. (2021), most involved a single school, district, or organizational unit, typically over one or two years. Only a few studies covered more than three years or engaged a substantial number of CI teams. Furthermore, nearly all the studies employed descriptive case study designs, offering rich qualitative insights but limited causal inference. Only three studies included comparison groups, and none used randomized controlled trials.

This methodological pattern underscores a central tension in the CI literature: While CI emphasizes contextual responsiveness and practitioner empowerment, these same characteristics make it difficult to evaluate using traditional experimental designs. The studies reviewed were primarily concerned with documenting processes and identifying perceived supports or barriers, rather than systematically measuring outcomes or establishing causal claims.

2.3 Culture of Evaluation and Capacity Building in Organisations

The establishment of a culture of evaluation is one of the fundamental pillars for guaranteeing the effectiveness, transparency, and sustainability of educational interventions. This approach encompasses more than a mere series of technical practices; rather, it signifies an institutional pledge to the principles of continuous learning, informed decision-making, and accountability for outcomes. As Mayne (2008) asserts, the establishment of an evaluative culture necessitates the transformation of evaluation from a discrete activity to a perpetual practice, seamlessly integrated into the processes of planning, implementation, and the refinement of actions.

This challenge assumes particular significance within the domains of *Non-Governmental Organisations* (NGOs) and schools, where evaluation is frequently perceived as an external imperative associated with funding or legal obligations. To overcome this instrumental perspective and cultivate a bona fide evaluation culture, it is imperative to nurture internal capacities in *Monitoring and Evaluation* (M&E capacity-building). Görgens and Kusek (2009) posit that the development of these competencies occurs across three distinct levels: the individual level, characterised by technical and analytical training; the organisational level, encompassing the integration of M&E systems and processes; and the institutional level, influenced by standards, incentives, and leadership that is conducive to evaluation. An effective M&E system must be appropriate to the context, flexible, participatory, and orientated towards the practical use of the data collected.

In the context of educational institutions, cultivating this evaluative culture entails the integration of institutional self-evaluation practices and the active involvement of diverse stakeholders (teachers, students, parents, staff) in the construction of indicators and the analysis of outcomes. This approach, frequently termed *participatory evaluation*, advocates for the involvement and empowerment of stakeholders, thereby reinforcing the legitimacy of the decisions made. Cousins and Whitmore (1998) seminal work established the foundations of this theoretical framework, positing that this type of evaluation is predicated on the co-construction of knowledge and the deliberate utilisation of evaluation processes as a conduit for organisational change.

Furthermore, the involvement of stakeholders facilitates the adaptation of performance indicators to local realities, thereby promoting enhanced congruence between strategic objectives and on-the-ground practices. Research conducted by Preskill and Boyle (2008) indicates that organisations that adopt a participatory and formative approach to evaluation tend to demonstrate superior long-term results, attributable to their augmented capacity for learning and adaptation. This logic is particularly suited to educational contexts, where impacts are often intangible, progressive, and influenced by multiple factors.

Another significant element pertains to the role of leadership in the consolidation of a *Culture of Evaluation*. School leaders or project coordinators who value evidence, allocate resources to evaluation, and integrate the results into decision-making cycles tend to generate environments that are more conducive to innovation and continuous improvement (Earl and Katz, 2002). It is imperative to emphasise that cultivating an evaluation culture is a protracted and non-linear process. To facilitate such a process, investment in training is required, as well as time for reflection,

methodological flexibility, and - above all - a strategic vision of evaluation as a tool for organisational transformation.

2.4 Towards Evidence-Based Educational Innovation

A growing emphasis is placed on the way data and impact results are presented and communicated. The increasing use of data visualisation through interactive dashboards, accessible infographics, and executive reports has been shown to make information more useful for several different stakeholders, including institutional decision-makers and those directly involved in schools. This democratisation of access to information has been shown to increase the involvement of school communities, thereby promoting a culture of transparency and co-responsibility.

It is also important to highlight the concrete benefits that impact measurement brings to the *institutional knowledge* (IK) of each project or organisation. Primarily, it facilitates systematic learning about the effectiveness of the methodologies applied, leading to their adaptation or reformulation based on evidence. Secondly, the longitudinal analysis of the data makes it possible to identify patterns of success, risk zones, or more vulnerable populations, enriching the capacity for strategic planning. The knowledge thus engendered contributes to the establishment of a robust organisational memory, supported by real data, which can be mobilised in funding application processes, partnerships, or dialogue with public bodies. Consequently, impact measurement is not merely an external requirement, but rather a fundamental management and internal development tool for organisations.

The importance of evaluating impact is therefore far from merely technical. It is a process that provides a framework for the reinforcement of the credibility and sustainability of educational initiatives. By demonstrating the value generated, it contributes to the legitimacy of the actions among stakeholders and creates the conditions for their replication on a larger scale. As authors such as Mayne (2010) have noted, the implementation of an effective evaluation culture has been shown to enhance not only the quality of interventions but also their social and political recognition.

Considering the context, the *Kaizen Education Movement* adopts an innovative approach by proposing an integrated strategy for continuous improvement, in conjunction with a systematic impact measurement model. By structuring pertinent indicators, developing data collection tools, and creating a logic of continuous feedback, the project seeks to strengthen the quality and efficiency of school pathways, while consolidating a solid and transferable institutional knowledge base. It is therefore the case that *Kaizen Education* responds not only to the current challenges facing education in Portugal, but also actively contributes to building a more informed, transparent, and results-oriented education ecosystem.

2.5 The Growing Culture of Impact Measurement in Portuguese Education

The measurement of the impact of educational projects in Portugal has become an essential practice for evaluating the effectiveness and efficiency of interventions in the education sector. While this approach is not yet fully entrenched across all initiatives, there is a growing commitment on the part of civil society organisations, foundations, and even public bodies to integrate more rigorous and systematised evaluation processes. This evolution has been driven by the need to produce evidence that supports strategic decisions, promotes continuous improvement, and justifies investments.

From a methodological perspective, a combination of quantitative and qualitative data collection methods has emerged. These include the analysis of school indicators, academic results, and attendance data, complemented by the collection of qualitative data through interviews, observations in educational settings, and focus groups with teachers, students, and parents. In certain more structured projects, quasi-experimental or statistical control methodologies, such as propensity score matching, are employed to compare similar groups that have been exposed or not exposed to the intervention. However, most initiatives continue to rely on descriptive analyses and case studies, which are more accessible but less robust in terms of inferences.

In relation to the indicators, a degree of convergence is observed across four primary domains: academic impact (e.g., enhanced grades or diminished retention), student wellbeing and engagement (e.g., levels of motivation, active participation, sense of belonging), transformation in teaching practices (e.g., extent of didactic innovation, utilisation of digital tools, teacher collaboration), and systemic impact or scalability (e.g., capacity for replication, geographical scope, influence on educational policies). It is noteworthy that these Key Performance Indicators (KPIs), although not universally defined across all projects, do tend to reflect the intended effects of the interventions in the short to medium term.

2.6 Impact Measurement Models and Approaches in a National Context

Impact measurement in educational projects in Portugal has evolved but continues to face significant challenges in terms of methodological rigour and practical application. This section presents an in-depth analysis of the models and approaches used in the national context, highlighting the main limitations and opportunities for improvement.

Quantitative vs. Qualitative Approaches

In Portugal, the focus of impact assessment in education has been predominantly centred on quantitative indicators, such as retention rates, grade point averages, and dropout rates. These indicators are frequently utilised to assess academic performance and to monitor student progress.

Nevertheless, this quantitative approach, while undoubtedly beneficial, may not fully capture the intricacies of educational experiences and the contextual factors that influence outcomes.

Conversely, qualitative approaches such as interviews, focus groups, and observations have received comparatively less attention, despite their capacity to yield profound insights into educational processes and stakeholder perceptions. The integration of qualitative methodologies has been shown to facilitate a more holistic and contextualised evaluation of the impacts of educational initiatives (Patton, 2002).

Use of Quasi-Experimental Designs

The utilisation of quasi-experimental designs in the context of education impact measurement in Portugal remains underdeveloped. The use of non-randomised control groups and the comparison of results before and after the intervention are pivotal in establishing causal relationships between interventions and observed outcomes. However, the scarcity of resources, lack of specialised training, and institutional resistance hinder the adoption of these more rigorous methods (Leite et al., 2023).

The absence of quasi-experimental designs compromises the ability to attribute observed changes directly to the interventions, limiting both the internal validity of conclusions and their generalisability.

2.6.1 Lack of Baselines, Control Groups, and Validated Scales

The absence of baseline data prior to the implementation of interventions poses a significant challenge to accurately measuring change. Additionally, the lack of control groups impairs the capacity to draw meaningful comparisons between populations exposed and unexposed to the intervention, thus weakening causal attributions.

Furthermore, the use of non-validated measurement instruments - or those adapted without rigorous methodology - raises concerns regarding the reliability and validity of collected data. Employing validated scales is essential for ensuring the accuracy, comparability, and robustness of impact assessments (Rossi et al., 2004; American Educational Research Association et al., 2014).

2.6.2 Current Limitations of Impact Measurement in Portugal

Despite the evolution of impact measurement practices, structural and cultural constraints continue to impede their potential to serve as tools for continuous improvement and evidence generation. A key issue is the limited technical capacity within educational institutions and NGOs to develop and sustain effective monitoring and evaluation (M&E) systems. This is evident in the scarcity of specialised personnel and the lack of ongoing training in evaluation methodologies, indicators, and data analysis.

Many schools and NGOs operate with small, overburdened teams, limiting their ability to collect and analyse meaningful data. Consequently, the utility of such data for pedagogical or strategic decision-making is reduced.

Additionally, an institutional preference for summative and external evaluations conducted at the end of project cycles often overshadows the value of formative, continuous assessment. This tendency frames evaluation as a final validation process, primarily aimed at reporting to funders or external audiences, rather than as an iterative, learning-oriented tool that can inform and adjust interventions in real-time.

Another recurring issue is the misalignment between project goals and the indicators used to measure success (Gertler et al., 2016). Frequently, selected indicators fail to reflect the intended impacts due to being overly generic, difficult to operationalise, or disconnected from school realities. This discrepancy weakens the evidentiary basis for assessing effectiveness and undermines efforts to improve or replicate interventions.

Lastly, many external evaluations are conducted with a predominantly communicational purpose - focused on public legitimisation rather than institutional learning. These evaluations often prioritise success stories and descriptive statistics over critical analysis of success factors, stakeholder involvement, contextual conditions, and causal mechanisms.

These challenges highlight the urgent need to cultivate a more mature and integrated evaluation culture in Portuguese education - one that values methodological credibility and practical relevance for all stakeholders.

Chapter 3

Problem Description

3.1 From Business Efficiency to Educational Excellence: The Evolution of the Kaizen Excellence in Education

The *Kaizen Business System* (KBS) has become a fundamental model for continuous improvement in organisations, especially in industries and companies, with a focus on the elimination of waste, the optimisation of processes and the promotion of a culture of sustainable innovation. This system, which combines efficiency, continuous improvement and innovation, was initially designed for business environments. However, it was soon realised that the principles behind it could also be applied in other contexts, such as education.

Education faces similar challenges to those encountered in business and industry. The necessity for continuous enhancement constitutes one of the fundamental pillars upon which educational institutions are founded. This necessity encompasses a wide range of activities, including the development and implementation of more effective curricula, as well as the effective management of human and financial resources. The constant search for quality and efficiency in teaching and learning processes requires a flexible and constantly evolving approach, as *Kaizen* proposes (Imai, 2012).

It was with this vision that the *Kaizen Excellence in Education* (KEE) was established, adapting the fundamental principles of *Kaizen* to the educational context. The core focus of KEE is to effect transformation in the domain of education, with the objective of enhancing the efficiency and sustainability of educational institutions, whilst concomitantly ensuring that they are oriented towards the continuous development of their students and teachers (as seen in 3.1). The application of the *Kaizen* philosophy has been demonstrated to facilitate the optimisation of teaching methods, the enhancement of educational management, and the establishment of an environment in which all stakeholders - including students, teachers, and administrators - are engaged in perpetual improvement.

A pivotal element of the KEE is its structure, which is predicated on continuous evaluation and improvement processes. This suggests that all members of the educational institution, including managers, teachers, students and even families, must be involved in a constant journey of learning

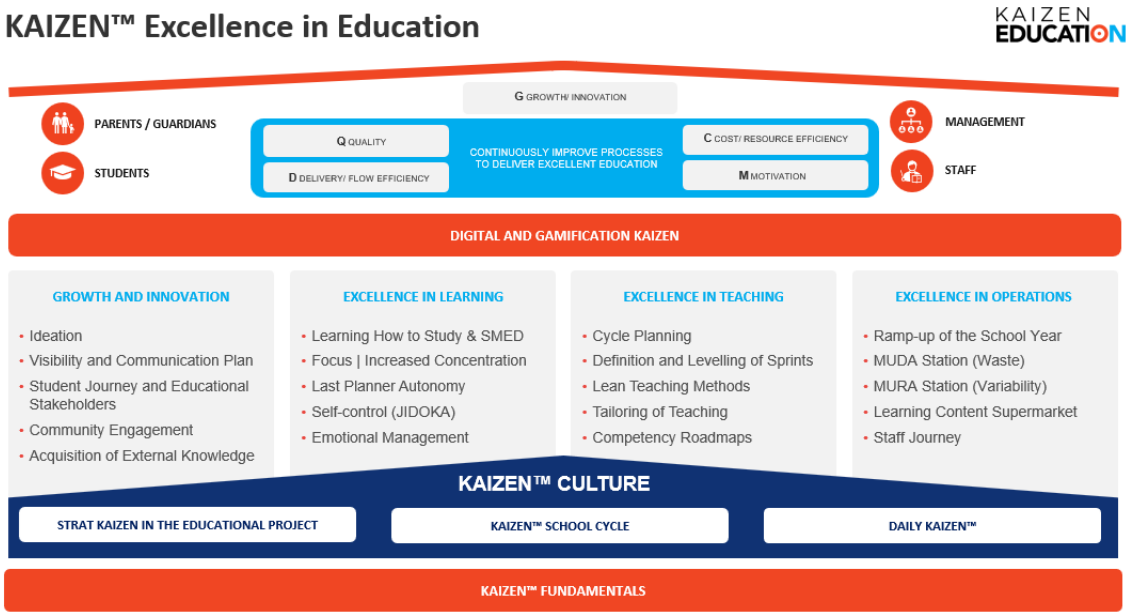


Figure 3.1: *Kaizen Excellence in Education* Roadmap

and growth. The central idea is that improvement is not a single goal, but a daily practice that applies to all levels and functions of the institution.

The fundamental principle of the *Kaizen Excellence in Education* is the promotion of a constant cycle of feedback and evaluation, which enables the continuous adaptation of educational strategies. This necessitates the implementation of continuous monitoring processes, which can be facilitated by tools such as visual management boards and the analysis of KPIs. These indicators are commonly utilised in the *Kaizen* environment and are now being adopted within the teaching and learning context. These Key Performance Indicators (KPIs) are not confined to the assessment of student performance, but also encompass the quality of teaching methods, administrative management, the resources utilised, and the interaction between all stakeholders.

Another fundamental aspect of the KEE is the integration of digital technology. Digitalisation in education, when aligned with *Kaizen* principles, has the potential to serve as a potent instrument for the optimisation of processes and the assurance of the most efficacious utilisation of resources. The utilisation of digital platforms for educational purposes, facilitating communication between educators and students, and enhancing school administration, presents a valuable opportunity to implement *Kaizen* in a more systematic manner. This approach not only streamlines information management but also enables more precise evaluation of educational processes.

In the field of education, the pursuit of continuous improvement is not confined to the development of innovative teaching methodologies. It also encompasses the establishment of an environment conducive to the flourishing of teachers and students within a perpetually evolving context. The importance of ongoing professional development for educators cannot be overstated, as it is instrumental in ensuring that they remain at the forefront of pedagogical and methodological advancements. In a similar vein, it is imperative that students are encouraged to adopt a

mindset of continuous improvement in their own learning process.

The KEE is not an isolated process, but rather a philosophy that is integrated into the running of the educational institution. Switching to an *Educational Kaizen Model* necessitates a cultural transformation, initiated by educational leaders but involving all individuals involved in the process. For this to be realised, it is essential that educational institutions develop participatory management models, whereby teachers, students and even parents play an active role in the identification of problems and the development of solutions. This collaborative approach enables all parties to share a common goal of continuous improvement, thereby increasing the system's chances of success.

Regarding organisational structure, the KEE has been designed to be flexible and adaptable, with the objective of allowing it to be adjusted to the specifics of each educational institution. It integrates concepts such as *Value Stream Kaizen*, which optimises the flow of information and resources within the institution, and *Daily Kaizen*, which focuses on the constant improvement of daily processes - from class management to student assessment. The adoption of a daily PDCA (Plan-Do-Check-Act) cycle, one of the main tools of *Kaizen*, is also crucial to ensure that changes are implemented effectively, and that progress is monitored continuously.

Finally, it is important to emphasise that the *Kaizen* philosophy, when applied to the KEE, is not restricted to academic performance alone. The system is holistic in nature, with the objective of creating an educational environment that is more efficient, innovative, inclusive and sustainable. It is intended to empower all those involved so that they can contribute to continuous and sustainable development. KEE was built to have the capacity to effect change within educational institutions, thereby facilitating the realisation of their academic objectives. Moreover, it has the potential to equip citizens with the necessary skills to effectively navigate the challenges of the contemporary world, fostering a perpetual spirit of continuous improvement.

3.2 The Importance of Impact Measurement in Social Projects

The evaluation of educational projects is of paramount importance in ensuring the effectiveness, sustainability, and relevance of interventions within the school context. In the context of social projects applied to education, this practice is an indispensable tool for understanding whether the changes introduced are effectively transformative and sustained over time.

The implementation of an intervention, in and of itself, does not guarantee the production of significant results. It is only through the systematic collection of evidence that the usefulness, efficiency, and relevance of the actions undertaken can be gauged. Contemporary impact assessment transcends mere technical obligation, representing an ethical dimension of responsibility towards target audiences, funders, and the educational community (Mayne, 2008).

In the specific context of education, this measurement facilitates the assessment of quantifiable academic outcomes, including performance progress and reduced absenteeism. Additionally, it enables the identification of less tangible changes, such as heightened motivation, enhanced school climate, and the consolidation of collaborative practices among teaching staff. The generation of

reliable data enables educational institutions and their collaborators to make informed decisions and modify their strategies in real time, thereby fostering a logic of adaptive management. As Mayne (2008) observe, the absence of an evaluation system hinders not only the measurement of real impact, but also the development of an institutional memory that fosters learning and guides future strategic decisions.

However, it is important to distinguish between two fundamental moments in this process: *continuous monitoring* and *impact assessment*. Monitoring is defined as a regular and systematic process of collecting and analysing data during project implementation. It is imperative that this monitoring is conducted to identify any discrepancies between the original plan and the actual results. This process enables the implementation of necessary adjustments and ensures the methodological coherence of the intervention.

The establishment of a feedback system is pivotal in fostering transparency and accountability throughout the intervention. This system is founded on the periodic administration of surveys to students, teachers and operational school assistants, the collection of process indicators (e.g. attendance at sessions, levels of participation or adherence to the proposed activities), and the convening of technical follow-up meetings. The efficacy of these monitoring mechanisms is particularly pronounced when associated with formal mid-term evaluations, which facilitate the timely collection of partial data and the subsequent informed decision-making process (OECD, 2022a).

Conversely, impact evaluation is conducted at a subsequent stage, with the objective of measuring the degree to which the delineated objectives have been realised, in addition to evaluating the enduring implications of the intervention on the school community. Monitoring is an activity which focuses on the process, whereas impact assessment is an activity which focuses on the results. To undertake a rigorous evaluation, it is imperative to establish a clearly defined baseline against which the final data can be evaluated.

The application of robust methodologies is also implied, encompassing quantitative approaches such as the analysis of grade or retention rate evolution, and qualitative approaches including interviews, focus groups and direct observation. As White and Sabarwal (2014) have observed, the rigour of impact assessment is essential to distinguish spurious correlations from causal effects; this is why the use of quasi-experimental models is becoming increasingly common in the field of education.

This distinction between monitoring and impact is not merely conceptual; rather, it reflects different operational needs. The former aims to optimise implementation and facilitate a rapid response to emerging constraints; the latter seeks to answer the central question of any intervention: what has changed, for whom, and why? As Gertler et al. (2016) observes, a significant challenge in educational evaluation is the tendency to focus exclusively on quantifying activities or outputs, while neglecting to undertake a critical analysis of their actual consequences. Impact measurement, when correctly structured, avoids this pitfall by refocusing on the results obtained and the underlying transformation processes.

In this sense, the importance of investing in solid and methodologically consistent evaluation systems is widely recognised, not only by academic literature, but also by international bodies such

as the OECD or UNESCO. The employment of pertinent indicators, adapted to the intervention context, in conjunction with the interrelation between quantitative and qualitative instruments, facilitates a more comprehensive and rigorous interpretation of reality. It is only through this approach that one can truly comprehend the social value of an educational project and ensure that the lessons learnt are mobilised to influence public policies and strengthen the school as a space for continuous improvement.

3.3 Current educational situation and its economic relevance

Recent decades have witnessed considerable progress in the Portuguese educational context, notably in the areas of reducing early school leaving rates, enhancing teacher qualifications, and ensuring universal access to basic and secondary education. Nevertheless, this progress has not been systematically accompanied by a culture of internal innovation in schools that would enable them to consolidate the dynamics of sustained continuous improvement. In terms of day-to-day practices, a pedagogical model that is strongly centred on curriculum reproduction and summative assessment prevails. In this model, the logic of meeting targets and programmes often supplants critical reflection on teaching and learning processes (Vincent-Lancrin et al., 2019).

Considering these observations, it is evident that there is a conspicuous absence of initiatives that systematically integrate continuous improvement concepts and tools into the Portuguese educational sector. It is evident that initiatives encompassing the active involvement of students and educators in the identification of problems, the experimentation with solutions, and the collective monitoring of outcomes are sporadic and not widely disseminated.

The absence of structured innovation is compounded by a significant gap: the inadequacy of impact measurement practices within the educational context. Despite the prevalence of projects with transformative intentions, there is a paucity of those that incorporate robust mechanisms for evaluating the effects produced from the outset. These effects may be categorised as either academic performance, student motivation, the well-being of the school community or the development of transversal skills. As the OECD (OECD, 2019a) emphasises, most educational initiatives in Portugal continue to be evaluated based on descriptive indicators or reports that fall short of the required standard, without the use of comparative methodologies, baselines or validated evaluation instruments.

This lack of methodological rigour also limits the broader contribution of education to sustainable development. Economist Eric Hanushek (Hanushek, 2005) demonstrates that the quality of education - measured through performance in international standardised tests such as PISA (The Programme for International Student Assessment) - has a substantially greater impact on economic growth than the mere number of years of schooling. His research, published by the International Monetary Fund (IMF), reveals that a one-standard-deviation improvement in test scores correlates with a 1% increase in annual GDP per capita growth.

Similarly, the OECD report *The High Cost of Low Educational Performance* (OECD, 2010) shows that even modest improvements in learning outcomes can lead to substantial gains in productivity and long-term economic performance. These findings underscore the strategic importance of not only expanding access to education but also enhancing its quality through mechanisms that ensure its effectiveness and continuous improvement.

The absence of a methodological framework has substantial ramifications. Firstly, it is challenging to recognise the value generated by interventions, which limits their replicability and scalability. Secondly, it hinders the accumulation of valuable evidence for policymakers and educational agents themselves, who frequently function without access to substantial data on effective practices, their application in specific contexts, and the underlying reasons for their efficacy. In conclusion, this phenomenon serves to diminish the capacity of academic research to engender applicable knowledge, thereby engendering a discrepancy between the practices undertaken within the field and the discourse that occurs within higher education institutions.

In this sense, the necessity for applied research in the domain of continuous improvement in education is evident. The development of scientific knowledge concerning intervention models, evaluation methodologies and sustained pedagogical management practices remains in its infancy in Portugal, particularly in comparison to other European countries that have a more extensive history of school improvement research.

The *Movement* thus aims to help fill this gap, offering not only a proposal for practical intervention, but also a methodological architecture that will allow its impact to be rigorously evaluated. It positions itself as a bridge between pedagogical innovation and the production of evidence, as part of a wider movement to valorise the school as a space for institutional learning, beyond the classroom - an ambition achieved by combining implementation with research.

3.4 The true value of Data

Data is increasingly recognized as the foundation of sound, evidence-based decision-making. When reliable and well-structured, data enables a deeper understanding of complex issues, allowing stakeholders to anticipate outcomes, streamline processes, and evaluate results with clarity. However, its usefulness is fundamentally tied to its quality. Poor data can mislead decision-makers, result in inefficient allocation of resources, and ultimately erode public trust in the systems designed to serve them.

This is particularly true in the context of social interventions, where decisions often have direct and lasting impacts on individuals and communities. Whether in education, healthcare, or social policy, high-quality data is essential not just for tracking progress, but for ensuring that support reaches those who need it most. Disaggregated, detailed data, especially at the individual level, enables more precise, equitable interventions that aggregated figures alone cannot support.

For example, in education, having access to student-level data reveals patterns and disparities that are otherwise hidden within class or school-wide averages. This level of detail is critical in designing interventions that are responsive to the real needs of learners. In public health, similarly,

accurate and timely data is vital for monitoring disease trends, targeting responses, and ensuring that no population is overlooked.

Beyond its practical applications, good data strengthens the integrity of social initiatives. It fosters transparency, enabling progress to be tracked and evaluated, while also building accountability among those responsible for delivering outcomes. This continuous feedback loop allows for necessary course corrections and promotes more effective use of resources.

Yet, collecting data at this level of granularity is not without challenges. It is often a time-consuming process that requires the cooperation of multiple institutions and careful attention to data protection and ethical considerations. Still, as Ghorna et al. (2024) underscore in their work, such efforts are indispensable. They note that "the quality of data used in these processes significantly impacts the accuracy, reliability, and relevance of insights derived", highlighting the central role of data integrity in meaningful analysis.

To serve its purpose, data must meet key criteria: it must be valid, reliable, timely, accurate, and ethically sourced. When these elements are lacking, the risks extend beyond poor performance - they can result in decisions that unintentionally perpetuate inequality or cause harm.

This level of data detail aligns directly with the core mission of the initiative: improving outcomes for each student, ensuring that no one is left behind. To date, this approach is unprecedented in education projects in Portugal, at least among those publicly available. The attention to data quality is, indeed, commendable.

3.5 Alignment with international recommendations

The measurement of impact in education, when aligned with the guidelines of international organisations, contributes significantly to the quality, usefulness and comparability of the data produced. In this regard, organisations such as the Organisation for Economic Co-operation and Development (OECD), UNESCO and the Education Endowment Foundation (EEF) have played a central role in defining benchmarks for the evaluation of evidence-based educational interventions.

The OECD has been emphasising the importance of integrated evaluation frameworks geared towards improving school results, both at an individual and institutional level. In the report entitled *Evaluation and Assessment Frameworks for Improving School Outcomes* (OECD, 2013), fundamental principles for effective evaluation are listed. These principles include the clear definition of educational goals, the careful selection of relevant indicators, the balance between quantitative and qualitative methods, and the structured return of the information collected to all educational agents.

The organisation also underscores the necessity to transition from a culture predicated on accountability and the measurement of summative outcomes, towards a formative approach that fosters teacher professional development and sustained pedagogical innovation.

Notwithstanding the existence of these reference frameworks, the Portuguese reality is still in a consolidation phase in terms of their systematic application. When compared to countries such as the Netherlands, Finland or the United Kingdom, where there is a consolidated tradition of

school improvement research and the systematic use of data for decision-making, Portugal reveals a more fragmented scenario.

The evaluation of impact continues to be predominantly reliant upon the initiative of civil society organisations or individual projects that receive external funding. To illustrate this point, the EEF advocates the implementation of rigorous scientific methodologies, including randomised controlled trials and systematic evidence reviews. The objective of this advocacy is to facilitate the development of pedagogical practices that are informed by empirical evidence and the proven impact of these methodologies. This approach remains somewhat uncommon in Portugal.

In this context, the *Kaizen Education Movement* demonstrates a deliberate effort to align with these international guidelines. The initiative proposes a continuous improvement model articulated with a structured impact measurement system, with the objective of integrating the six evaluation criteria defined by the OECD Development Assistance Committee (OECD-DAC): relevance, coherence, effectiveness, efficiency, impact and sustainability (OECD, 2019b). The definition of relevant indicators, the regular and systematised collection of data, the involvement of multiple stakeholders, and the use of information for learning and internal planning are all consistent with the best practices promoted by international bodies.

The evaluation design incorporated a simultaneous control group during the implementation phase in the Agrupamento de Escolas do Cerco do Porto, enabling direct comparison between intervention and non-intervention cohorts. To enrich this perspective, a simulated historical comparison was also employed, allowing for a broader understanding of changes over time, particularly in the absence of pre-intervention baseline data. This dual strategy reflects best practices in educational evaluation, combining methodological rigor with contextual adaptability.

This strategic alignment serves to enhance the project's scientific and institutional credibility, while concurrently enabling its integration into a broader discourse concerning European and international debates on innovation and educational effectiveness. This approach contributes to the consolidation of a national evaluation culture that is anchored in empirical evidence and oriented towards the continuous enhancement of educational processes.

Chapter 4

Solution

4.1 Rethinking Impact Measurement: From Institutional Metrics to Student-Level Insights

A recurring limitation in impact measurement models used in Portugal is their predominant focus on aggregated data at the school or classroom level, often across multiple years or institutional cohorts. Evaluations tend to rely on macro indicators such as school performance averages, retention rates, or standardized test results to assess the efficacy of educational interventions (Santiago et al., 2012; Fernandes, 2009). While these metrics are valuable for capturing system-wide trends, they often obscure individual-level dynamics, failing to account for the nuanced trajectories of students within different learning environments. Crucially, student-level data - such as attendance, grade progression, assessment results, or behavioral indicators - are rarely integrated into impact analyses, even though they offer a far more granular and actionable understanding of learning processes and outcomes.

One of the most pressing challenges, therefore, lies in shifting the unit of analysis downward - from institutions to individuals. This requires not only the collection of fine-grained, student-specific data, but also the capacity to process, anonymize, and interpret this data responsibly and ethically. By grounding impact measurement in student-level information, educational programs can better capture differential effects, identify targeted areas for intervention, and generate feedback loops that meaningfully support teachers and learners alike.

To achieve this paradigm shift, two core conditions must be addressed. First, it is necessary to develop or gain access to reliable and scalable data infrastructures capable of integrating student-level variables over time. These might include national *Education Management Information Systems* (EMIS), school-based digital platforms, or custom-built longitudinal databases that link academic, behavioral, and socio-emotional indicators. Second, evaluators must adopt or develop methodologies that are suitable for analyzing individual-level impact, such as student growth percentiles, value-added models, or quasi-experimental micro-level designs (Angrist and Pischke, 2009).

Such a shift aligns with international calls to make educational evaluation more learner-centered and equity-informed. By focusing on the "base of the pyramid" - the student - it becomes possible to trace how and why certain learners benefit from interventions while others do not, and to adjust pedagogical approaches accordingly. This reconceptualization of impact measurement is not merely a technical refinement; it is a strategic commitment to educational justice, ensuring that system-level improvements translate into real gains for the learners who matter most.

4.2 Definition of the Impact Model

The ideal scenario for this project envisions the development of a validated impact measurement model, one that is highly adaptable to the diverse realities of schools and supported by clear, objective, and easily monitorable indicators. The effectiveness and sustainability of such a model would depend on the active and ongoing involvement of all key stakeholders - including schools, policymakers, foundations, and local communities - within a framework of shared responsibility and transparent data practices.

Moreover, this scenario assumes the availability of adequate technological and human resources, capable of supporting not only the effective implementation of the model but also its progressive scaling across the national education system. Over time, the model could exert a direct influence on the design and revision of public education policies, contributing to a system that is more responsive, equitable, and student-centered.

However, there is also the potential for a more limited scenario, where the implementation of the model is restricted to pilot schools. This would allow for the achievement of positive results but limit the extent to which these results can be disseminated. The model may be operationalised with the establishment of dashboards and the definition of KPIs, but the conditions for immediate scaling up may not be in place. Stakeholder involvement may be partial, with the need to strengthen partnerships throughout the process, and replication of the model in other schools will depend on the availability of resources and political will.

Additionally, a set of auxiliary indicators was calculated specifically for this analysis, providing complementary insights to support the core evaluation metrics:

- Total cost of retentions per cycle of studies in public education: Based on national educational statistics, the total cost of student retentions was estimated for each cycle of public education. This indicator reflects the overall financial impact on the education system resulting from students repeating a grade, accounting for additional teaching resources, administrative efforts, and extended use of infrastructure. It highlights how grade retention contributes to increased public expenditure and inefficiencies across the educational system.

- Estimated cost of one class-period absence per student: To further quantify the impact of absenteeism, the cost associated with a single class-period absence per student was estimated. This calculation supports a more detailed understanding of how absenteeism contributes to inefficiencies and financial strain within the education system.

- Average income gain per student from an additional year of schooling: The research also estimated the individual economic return on education by calculating the average increase in annual income that results from one additional year of schooling. This indicator provides insight into the long-term personal financial benefits of educational attainment and underscores the broader societal value of promoting school completion. It serves as a key argument for policies that reduce early school leaving and encourage progression through the education system.

These indicators, along with the Social Return on Investment (SROI) framework, allowed for a more nuanced analysis of the impact of *Kaizen* principles on education. They provided a tangible way of quantifying the social and financial benefits of continuous improvement in schools, linking educational improvements to both short-term academic performance and long-term socioeconomic outcomes.

This model distinguishes itself through its degree of specificity at the individual student level, a key differentiator from other educational projects implemented in Portugal. As said before, while most national initiatives and evaluations rely on aggregated indicators at the class, school, or municipal level, this model places the student at the centre of the analysis, enabling granular tracking of academic progress, engagement, and behaviour. This individualized focus allows for earlier, more precise interventions and aligns with a pedagogical vision that is responsive to student needs and differentiated learning pathways.

Nevertheless, gathering reliable and comparable student-level data across schools presents real challenges. Tracking absences and behavioural records proves difficult due to variation in data management systems and submission practices, which differ not only between school groups but also across platforms within the same region. This fragmentation limits the possibility of robust comparisons and highlights the need for a standardized, interoperable data infrastructure capable of integrating inputs from diverse school environments.

Moreover, the project highlights the importance of developing a tool to track classroom-level evolution over time, both between classes (e.g., *Kaizen* vs. non-*Kaizen* Control groups) and within the same class longitudinally. This includes establishing historical baselines, enabling future comparisons as more classes adopt the intervention - potentially eliminating natural control groups. In this scenario, comparing current class outcomes with cohorts from previous years (e.g., five-year historical trends) will be essential for understanding real improvement versus contextual fluctuation. Such a tool would provide decision-makers and educators with deeper insight into the model's sustained impact and replicability across cycles and contexts.

4.2.1 Implementation Potential and Systemic Influence of the Impact Measurement Model

The pilot implementation of this impact measurement model has demonstrated strong potential for improving educational outcomes and informing policy development across three key dimensions: academic performance, engagement and scalability (Education, 2025).

First, academic impact has been evident through observed improvements in student performance indicators. Preliminary data from participating schools show reductions in the number of failed subjects and increased year-on-year student progression rates. In schools such as Agrupamento de Escolas do Cerco do Porto, the percentage of students who reduced the number of negative grades over the academic year and improved their final average was significantly higher among those involved in *Kaizen*-supported interventions (see Chapter 5, *Results*).

Second, in terms of engagement, both students and educators reported high levels of motivation. This was captured through tools such as the Net Promoter Score (NPS), which showed a high likelihood of teachers recommending the initiative to peers, and similarly high satisfaction among students (Appendix D) - a critical indicator of the initiative's sustainability (Escola do Falcão, 2022).

Third, the scalability of the model has been validated by the growing number of students, schools, and municipalities involved in the pilot. The organic adoption of methodologies by schools outside the formal network further reflects the adaptability and perceived value of the approach. Since 2020/2021, the initiative has expanded to 8 school clusters across 4 different districts and is projected to reach over 2,000 students in 2024/2025.

4.2.2 From Pilot to National Framework: Cost-Based Rationale

A core component of this project is its potential to generate cost-efficient strategies by addressing one of Portugal's long-standing educational challenges: student retention. Drawing on data from EduStat (2022) and national budget figures, the financial cost of student retention in the public system has been estimated (DGEEC, 2023):

- Primary and Lower Basic:
  - 1st cycle retention cost: €41,074,413 (1.8% × 335,674 students × €6,798/student)
  - 2nd cycle retention cost: €38,869,570 (3.1% × 184,445 students × €6,798/student)
- Upper Basic and Secondary:
  - 3rd cycle retention cost: €160,850,343 (6.2% × 303,363 students × €8,552/student)
  - Secondary retention cost: €263,258,060 (9.8% × 296,371 students × €9,064/student)

In total, student retention across all cycles cost the public education system approximately €503 million in 2022, accounting for nearly 5% of the national education budget, which was approximately €10.5 billion (DGEEC, 2023). These figures underscore the urgent need for effective intervention models capable of reducing school failure and dropout rates. They also highlight how targeted educational improvements at the individual level can yield substantial system-level economic returns. This approach offers a more precise understanding of intervention effectiveness and allows for early identification of at-risk students, enabling timely and context-appropriate support.

4.2.3 Toward Scalable, Policy-Aligned Implementation

The long-term success of the model depends on its integration into national education strategies, supported by the availability of technological and human resources. Adequate investment in teacher training, data systems, and localized facilitation is critical to expanding the initiative sustainably. The ideal scenario envisions a validated, adaptable model, supported by clear indicators and the active engagement of schools, policymakers, communities, and philanthropic organizations. Its aim is to embed continuous improvement practices across the system, aligned with the OECD's call for coherence between school-level innovations and national policy frameworks (Santiago et al., 2012).

In contrast, a more constrained rollout - limited to pilot schools or municipalities - may yield localized successes but limit broader influence. In such a scenario, dashboards and KPIs may be developed, yet the absence of structural conditions such as funding, interoperability, and political alignment could hinder replication and scalability.

Nonetheless, by anchoring impact measurement in student-level outcomes and by clearly demonstrating both academic gains and cost-efficiency, this model presents a scalable and policy-relevant approach to educational reform. It offers a replicable framework that is not only data-driven and outcomes-focused but also grounded in equity, transparency, and continuous evaluation, principles essential for building resilient and inclusive education systems.

4.2.4 The Impact Measurement Approach

A critical step in designing any educational impact measurement model lies in clearly defining the comparative strategy to be adopted. Within the context of this project, three main approaches can be considered:

- Comparison with a Simultaneous Control Group (Intervention vs. Non-Intervention): This approach enables a more precise isolation of intervention effects by comparing classes or schools with similar sociodemographic, academic, and contextual characteristics, where only some receive the model's support. It aligns with the logic of Counterfactual Impact Evaluation (CIE), in which outcomes observed in non-intervention groups serve as an estimate of what would have happened to the intervention group in the absence of the program. This is a robust strategy, particularly useful during the initial years of implementation when some groups remain unexposed to the intervention (Shadish et al., 2002).

- Comparison with Historical Data of the Same Group (Longitudinal Evolution): When the majority of classes within a school or district have already adopted the intervention, maintaining "pure" control groups becomes difficult. In such cases, comparing current performance with historical results of the same class or previous cohorts with similar profiles gains relevance (Cook and Campbell, 1979). This approach measures performance evolution over time and is especially useful in multi-year cycles, such as the middle and upper levels of basic education (Durlak and DuPre, 2008).

- Continuous Monitoring Over a Complete Cycle: This strategy aims to assess the cumulative impact of the intervention over several academic years (e.g., from 1st to 4th grade), capturing both immediate improvements and sustained effects (Fullan, 2016). It is particularly relevant for understanding the maturation effect of methodologies based on continuous improvement and socio-emotional development (Bryk et al., 2015).

Choosing the most appropriate approach should therefore depend on the project's maturity stage, data availability, and the structure of the educational context. In initial phases, comparison with simultaneous control groups tends to be more effective and informative. In more advanced stages, with broader project coverage, historical and longitudinal analyses become essential to ensure ongoing evaluation and identify systemic impact trends, as Kane and Staiger (2012) concluded.

For the analysis of this project, the impact measurement was primarily conducted using a simultaneous control group approach, specifically within the *Agrupamento de Escolas do Cerco do Porto*, where the intervention was implemented in some classes while others served as controls. Additionally, historical data comparisons were simulated to complement the analysis, providing a longitudinal perspective despite the limited availability of real past data.

4.2.5 Selecting the Right KPIs

In the realm of educational transformation, where outcomes are deeply human, multifaceted, and long-term, selecting the right Key Performance Indicators (KPIs) becomes a critical foundation for meaningful impact assessment. For *Kaizen Education*, which is rooted in principles of continuous improvement and learner-centered development, KPIs must do more than quantify: they must reflect learning progress, equity, engagement, and systemic growth. Poorly chosen KPIs can undermine even the most well-intentioned efforts, leading to superficial success while missing deeper educational outcomes.

The power of KPIs lies in their alignment with purpose. When indicators are selected based on convenience or tradition rather than strategic and pedagogical relevance, they risk distorting priorities. For instance, measuring only standardized test scores may ignore improvements in critical thinking, creativity, teacher capacity, or student well-being. Such a narrow focus may inadvertently promote teaching to the test or exclude vulnerable learners from the full benefits of a holistic education.

In the context of education, this challenge underscores the risk of prioritizing academic attainment metrics while neglecting equally important but less tangible outcomes such as inclusion, resilience, and long-term learning. This concern is echoed in the broader performance measurement literature. For instance, Taticchi et al. (2013), in their review of sustainability-related KPIs, highlight that many performance systems focus disproportionately on easily measurable economic outputs, often at the expense of environmental and social dimensions.

A similar perspective is offered by Neely et al. (2002) in their Performance Prism framework, which proposes that KPIs should be shaped by the needs and expectations of multiple stakeholders.

In the case of *Kaizen Education*, these stakeholders include not only students and teachers but also parents, policymakers, community members, and future employers. KPIs that fail to incorporate these perspectives risk focusing on isolated aspects of performance while missing the collaborative nature of educational success.

Developing such inclusive and strategic KPIs is not a simple task. Parmenter (2015) emphasizes that many organizations adopt "vanity metrics" that look good in reports but are disconnected from real improvement. For an education initiative, this could include metrics like enrollment numbers or completion rates that do not reflect quality, depth of understanding, or transformative outcomes. Parmenter advocates for rigorous processes that link KPIs to core success factors - such as student engagement, instructional quality, and learning continuity - all essential for a model like *Kaizen* that values lifelong learning.

As *Kaizen Education* advances, its performance indicators must evolve to reflect the organization's changing maturity and strategic ambitions. Bititci et al. (2015) emphasize that KPIs need to be aligned with both the current stage of development and the goals being pursued. In the early phases, simple measures such as attendance can provide useful insights, but as the system matures, it becomes essential to adopt more sophisticated metrics. These should capture deeper, longer-term outcomes like sustained learner progression, adaptability, and learner agency—qualities that resonate strongly with *Kaizen*'s overarching vision for continuous improvement.

The importance of balanced and forward-looking indicators is also emphasized in Kaplan and Norton's Balanced Scorecard (Kaplan and Norton, 1996). While originally developed for business, its principles are highly relevant to education. The framework suggests that performance measurement should reflect not only outcomes but also the processes, relationships, and capacities that drive them. In *Kaizen Education*, this means measuring both what learners achieve and how learning is supported - through pedagogy, teacher development, community involvement, and adaptive learning environments.

Moreover, Ghalayini and Noble (1996) stress the danger of static, cost-based performance systems that fail to capture agility, creativity, and responsiveness - qualities deeply relevant to modern education. Their critique reminds us that KPIs must reflect the dynamic, evolving nature of educational ecosystems. In *Kaizen*'s model, which prioritizes iterative improvement and learner empowerment, static indicators will fall short.

Franco-Santos et al. (2007) reinforce this need for alignment between measurement and decision-making. For Kaizen, this means ensuring that data informs both classroom practice and strategic planning, with feedback loops that help educators and leaders adapt continuously.

Finally, de Waal (2003) draws attention to the behavioral impact of performance metrics. In education, how something is measured can shape what is taught, how teachers feel about their work, and how learners experience school. If KPIs reward compliance over creativity or speed over understanding, they can distort the very culture that *Kaizen* aims to cultivate. Effective KPI systems must therefore be grounded not only in strategy but in values: trust, collaboration, equity, and growth.

4.3 Kaizen Education Impact Assessment: Applying the Model

To rigorously quantify the longitudinal effects of the *Kaizen Education* Model, a robust, data-driven impact assessment strategy was developed. This section outlines the rationale, methodological decisions, and analytical procedures that supported the evaluation of four academic years of intervention, offering a replicable framework for future applications and providing actionable insights to stakeholders.

The impact assessment was constructed based on raw institutional data extracted from a centralized relational database comprising over 200 separate tables. These tables contained highly granular information on academic performance, attendance, behavioral incidents, and enrollment patterns. Figure 4.1 illustrates the flow of data throughout its extraction, processing, and presentation.

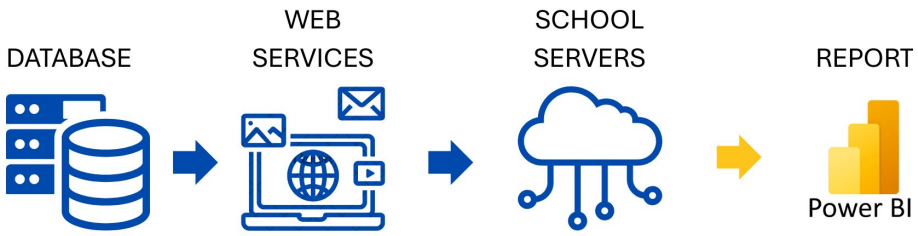


Figure 4.1: Flow of data from extraction to presentation

4.3.1 Methodological Design and Rationale

Given the complexity and volume of data, a phased and systematic approach was adopted. The initial phase involved identifying relevant tables through exploratory data analysis, guided by the central research question:

> *How has the implementation of the Kaizen model influenced student outcomes over time?*

One of the model's most distinctive methodological decisions was to use student-level data, as opposed to aggregated class or school metrics, as the unit of analysis. This choice was grounded in two key premises:

1. Accuracy in Attribution: Aggregated data often masks significant variations within cohorts, making it difficult to trace specific beneficiaries or mechanisms of impact. Student-level data enables fine-grained attribution of progress or setbacks to particular interventions.

2. Continuity and Comparability: Class compositions vary year to year due to enrollment shifts, school transfers, and administrative changes. Tracking class labels (e.g., 1A in 2022/23 and 2A in 2023/24) does not guarantee continuity. By isolating only students present across both years, the model ensures reliable, longitudinal comparisons.

4.3.2 Limitations and Adaptive Solutions

Several limitations emerged during implementation, significantly shaping the final model design:

- Data Entry Errors: Inconsistencies such as incorrect school IDs, ambiguous absence codes, and incomplete student profiles necessitated a rigorous data-cleaning phase. Missing fields were flagged, typologies standardized, and a data quality score was introduced.

- Terminological Inconsistencies: Variability in terminology across schools and years (e.g., use of "semester" vs. "period") prompted the development of a semantic harmonization layer to align variable names within a common framework.

- Limited Familiarity with Impact Measurement: At *Kaizen Education*, there was initially limited internal expertise in areas such as longitudinal analysis and control group design. This required a dedicated phase of research and self-education to build the necessary knowledge base, ensuring that the impact model was methodologically sound and data practices were robust going forward.

- Lack of Predefined Indicators: With no existing educational KPI framework in place, indicators were developed from scratch. These were aligned with national curriculum priorities and made comparable with international metrics (e.g., OECD standards).

These challenges required not only technical fixes but strategic choices about the assumptions embedded in the model. Ultimately, they highlighted the importance of building a flexible yet standardized analytical structure, one capable of adapting to local variability while maintaining systemic consistency.

4.3.3 Assumptions

For a measurement exercise of this scale, the analysis must be grounded in a series of fundamental assumptions. These assumptions guide analytical choices, establish validity criteria, and ensure the internal coherence of the model:

1. It is possible to create valid and comparable educational indicators even in the absence of a pre-existing framework.

   Although a consolidated framework facilitates the definition and alignment of indicators, its absence does not prevent the construction of reliable and useful metrics for evaluating educational policies. According to the OECD (2022a), the process of selecting indicators should be guided by a combination of clear objectives, local relevance, and international standards. This ensures that metrics reflect contextual priorities while enabling comparability with other systems and experiences.

2. Student-level data analysis enables a more accurate and realistic assessment of educational impact over time.

   The choice to use individual-level data-rather than class- or school-level aggregates represents a significant methodological advancement. This approach provides greater granularity, allowing for the detection of subtle variations between students and control for individual-level factors often obscured in aggregated data. Additionally, it permits sample isolation by excluding students who enter or leave a class between years.

   Sakız et al. (2021) demonstrate the benefits of longitudinal student-level analysis in tracking learning determinants over time in higher education, reinforcing this approach as both statistically robust and centered on students' real experiences.

3. Methodological standardization can coexist with local adaptation, enabling model replication without sacrificing contextual sensitivity.

   As noted by the OECD (2022a), educational models can be designed with common analytical cores and flexible operational elements. This approach allows comparability across regions while accommodating local diversity. In the *Kaizen* model, for example, student-level data analysis was standardized, while terminology and tools were adapted to local practices-ensuring both technical consistency and contextual relevance.

4. Data cleaning is a foundational step in the research process, critical for ensuring the accuracy and reliability of analytical outcomes.

   Van den Broeck et al. (2005) emphasize that raw data often contains errors and inconsistencies that, if uncorrected, can compromise analytical validity. Effective data cleaning includes detecting, diagnosing, and correcting anomalies, as well as documenting procedures transparently to support reproducibility and credibility.

5. School records are sufficiently reliable for longitudinal analyses.

   The use of administrative school data is increasingly recognized in educational research due to its comprehensive coverage, temporal resolution, and low cost. Despite challenges such as inconsistent entries or assessment variations, recent research validates the reliability of these records when proper cleaning and harmonization techniques are applied.

Reardon (2019) used administrative data from over 45 million students across 11,000 U.S. school districts to explore variations in educational opportunity, showing that school records can serve as a robust foundation for large-scale longitudinal analysis. With proper treatment, administrative records support reliable impact evaluations and inform evidence-based educational policy.

4.3.4 Filtering Criteria and Scope of Analysis

The analysis focused exclusively on primary-level classes (Years 1 to 4) within the *Agrupamento de Escolas do Cerco do Porto*, spanning five consecutive academic years (2020/21 to 2024/25). This temporal range enabled the observation of multi-year trends and accommodated variations in implementation maturity.

To facilitate disaggregation and ensure analytical flexibility, data filters were created based on:

- Academic Year.
- Year Group (1st to 4th).
- School Unit (within the cluster).

This allowed both vertical (year-on-year) and horizontal (cross-sectional) comparisons, thereby supporting more nuanced and granular conclusions. One of the defining strengths of the model is its capacity to track individual student trajectories; This approach was essential to ensure that comparisons across time periods were not distorted by class composition changes resulting from student mobility, transfers, or dropouts.

The comparative framework was structured around two overarching conditions:

- *Kaizen* Classes: Groups that received direct support and structured interventions under the *Kaizen* model in the selected academic year.
- Control Classes: Equivalent groups from the same schools or clusters, without exposure to *Kaizen* interventions during the same timeframe.

Both groups were further segmented into:

- Year 0 (Pre-Intervention): Serving as the baseline.
- Year 1 (Implementation Year): Representing the active phase of the *Kaizen* implementation.

A deliberate distinction was made between:

- Year 0 Control: Never exposed to *Kaizen* before or during the analysis period.
- Year 0 *Kaizen*: Not yet exposed in the baseline year but scheduled for implementation in Year 1.

This distinction is methodologically significant. While neither group received the intervention during Year 0, the anticipation or institutional preparation in *Kaizen*-designated classrooms may already influence behaviour or expectations - a phenomenon described in the literature as *anticipatory effects* (Ariel et al., 2021). Ignoring this could lead to false assumptions about baseline neutrality.

The academic indicators chosen for evaluation included:

- Portuguese.
- Portuguese as a Non-Native Language.
- Mathematics.
- Environmental Studies.

Subjects such as English and Physical Education were excluded due to inconsistent evaluation criteria, elective status, or high variability in instructional time, all of which undermine comparability.

For each student, both baseline scores (Year 0) and outcome scores (Year 1) were collected. The individual-level change in score was then calculated, and average variations were computed for each subgroup (*Kaizen* and Control). The final impact measure was derived from the difference in these variations. This method enabled the estimation of the relative effect of the intervention while controlling for broader trends affecting all students.

Results were presented through a combination of interactive dashboards and analytical summaries, designed to make complex findings accessible to diverse audiences, including school leaders, funders, and policymakers. Special attention was given to visualising:

- Longitudinal evolution for common students.
- Comparative trajectories between Kaizen and Control groups.
- Subject-specific performance gains.
- Cohort-level trends by year group and school.

These visual tools serve both as instruments for internal pedagogical reflection and as decision-support mechanisms for external stakeholders. They enhance the transparency, legitimacy, and scalability of the *Kaizen* model.

4.3.5 KPIs Calculation Logic

As previously discussed, the accuracy and specificity of the results depend on a variety of contextual and methodological factors. In the case of primary-level students (Years 1 to 4) from the *Agrupamento de Escolas do Cerco do Porto*, the selected Key Performance Indicators (KPIs) were:

- The mean grade in core subjects.
- The percentage of failing grades, calculated as the proportion of assessments below the pass threshold.
- The average number of absences per student.

These KPIs are essential for evaluating multiple dimensions of educational quality. They enable not only the assessment of academic progress and teaching effectiveness but also provide insights into student engagement and attendance patterns. The latter is especially relevant, as absenteeism is often both a symptom and a predictor of broader academic challenges. In fact, consistent attendance is commonly regarded as a precondition for sustained academic success, and chronic absenteeism can reflect systemic inefficiencies and social vulnerability-thereby translating into a loss of public investment.

While the retention rate (i.e., the rate at which students are held back) is also part of the broader impact assessment model, it is not included in this specific analysis. This omission reflects the context-specific nature of the primary education cycle, where retention is minimal and not a widely adopted performance criterion. Nevertheless, the model is computationally prepared to incorporate this KPI for higher educational levels in future analyses.

The process of calculating KPIs follows a rigorous multi-step approach. First, student identifiers are selected based on class grouping. To enable year-on-year comparisons, students present in both the baseline year (Year 0) and the implementation year (Year 1) are matched using dataset intersections. Once this longitudinal cohort is established, the KPIs are computed by applying the filtering criteria described in the previous section, including subject, school unit, and semester:

- Mean grades and failure percentages are derived from students' numeric scores in the core subjects.
- Absences are calculated on a annual basis and reflect individual attendance records.

The following section presents the final dashboard solution, developed to serve as the primary visualisation tool for monitoring and communicating key performance indicators.

4.3.6 Report and Monitoring Charts

To ensure clarity and accessibility in communicating findings, an interactive dashboard was developed using Power BI. This platform allows dynamic visualisation of key metrics and analytical insights, tailored to the needs of diverse stakeholders-including funders, school management, and teaching staff. Each group can apply filters and explore data through lenses relevant to their specific responsibilities and goals.

Additionally, Appendix B (Figure B.1 and B.2) contains the prototypes of the dashboard, which guided the graphic design and report structure. This iterative development ensured early

alignment with project objectives and promoted clarity, user-friendliness, and data-driven decision-making.

The dashboard is organised into three main sections, each focusing on a different axis of analysis to provide a comprehensive, segmented view of the data:

- Analysis and annual comparison (last 4 years): This section presents a comparison between Year 0 (before the intervention) and Year 1 (after the intervention) over the last five years. The analysis is broken down by school and year of schooling, enabling patterns, trends and variations in academic performance to be identified in different contexts, from 2021/2022 to 2023/2024 (See Figure B.3 from Appendix B).

- Analysis of the First Semester (e.g., 2024/2025): This section presents a detailed analysis of the first semester results, using the 2024/2025 school year as an example. By directly comparing performance with the same period in the previous year, we can effectively monitor the early impact of the intervention in near real time. This allows for timely adjustments to strategies as needed (See Figure B.4 from Appendix B).

- This final section provides a longitudinal overview of the past five years, comparing the performance of *Kaizen* classes in the selected year with that of *Kaizen* classes separated by five years. The analysis aims to identify the programme's sustained impact and validate the intervention's effectiveness based on empirical evidence (See Figure B.5 from Appendix B).

Chapter 5

Results

To validate the entire thesis developed so far, the results calculated over the duration of the *Kaizen Education Movement* project at the Agrupamento de Escolas do Cerco do Porto are presented, more specifically in the primary schools of Corujeira, Falcão, Nossa Senhora da Campanhã, São Roque da Lameira, Lagarteiro, and Cerco.

These results represent the culmination of four years of implementation and dedication by all involved. As outlined in the previous sections, the selected indicators are average grades, percentage of failing marks (both measured at the end of the school year), and total absences throughout the year.

For example, the average grades directly reflect academic progress and serve as a key indicator to assess whether the pedagogical strategies implemented are effectively supporting student learning. A consistent increase in the average grades would indicate a positive effect of the movement's practices, whereas stagnation or decline would signal the need for adjustments.

The percentage of failing grades complements this analysis by revealing the proportion of students who did not meet the minimum required standards, functioning as a critical indicator for identifying groups at risk of failure. A reduction in this value over time would suggest an effective intervention contributing to the reduction of educational inequalities.

The total number of absences, in turn, is an important indicator of student engagement and motivation, reflecting aspects related to attendance and school commitment. Decreases in absences may indicate higher student involvement and an improved school environment, factors that positively influence academic success.

This evaluation reinforces the fundamental role of the *Kaizen Education Movement* as an initiative capable of promoting significant improvements in students' performance and attendance, aligning with the project's goals of inclusion and educational excellence.

5.1 Aggregated Results in AE Cerco do Porto

The sample used for this analysis consisted of 215 students who had participated in the *Kaizen Education* AE Cerco do Porto project continuously across consecutive academic years. Only

students who were enrolled in the program before and throughout the implementation period were included in the analysis, ensuring that the observed outcomes could be attributed more confidently to the program's long-term influence. While other students joined the classes during the year of implementation and may have also benefited from the intervention, they were not considered in the SROI calculations due to the lack of comparable baseline data.

The table below (Table 5.1) presents a summary of the results for each indicator. The calculations were conducted using two distinct approaches:

- School year analysis: All classes from the same grade (e.g., 2nd grade, 3rd grade, 4th grade) are grouped and analyzed together, regardless of the year they were implemented. For example, all 1st grade classes that received the intervention - whether in 2021/2022 or 2023/2024 - are analyzed as one group. (Table A.1, Table A.2 and Table A.3 of Appendix A).

- Academic year analysis: This approach groups all classes (from different grades) that were active during the same school year. For instance, results from all grades in 2022/2023 are aggregated and compared with 2023/2024. This approach provides a longitudinal view of the program's overall impact (also in Table A.1, Table A.2 and Table A.3 of Appendix A).

Key findings include:

- An increase in students' average grades in 83% of the cases (each unit of analysis corresponds to a class). On average, this improvement amounted to a 70% increase when analyzed by grade level and 39% when analyzed by academic year.

- A reduction in the percentage of failing grades in 100% of the cases. This reduction averaged 43% in the grade-level analysis and 35% in the academic-year analysis.

- A reduction in the number of absences in 83% of the cases, with an average decrease of 73% observed consistently across both analytical categories (details in Table A.3, A.5, A.6 and A.7 of Appendix A).

Important Considerations

While the results presented here are aggregated, it is important to note that the data can be disaggregated by school, grade level, and even classroom, depending on the specific interests of stakeholders.

It is also important to clarify that the figures discussed do not include data from the 2024/2025 academic year. These more recent results are presented in Table A.1, Table A.2 and Table A.3 of Appendix A and indicate a marked decline in average grades. However, this downward trend may be partially attributed to the increasing difficulty of the curriculum at the end of the academic year. It is common for student performance to improve progressively throughout the year; thus, comparing mid-year or first-semester results with those from the end of the year can distort the true impact of the intervention. For this reason, end-of-year or second-semester results offer a more reliable basis for evaluation and are prioritized in this analysis.

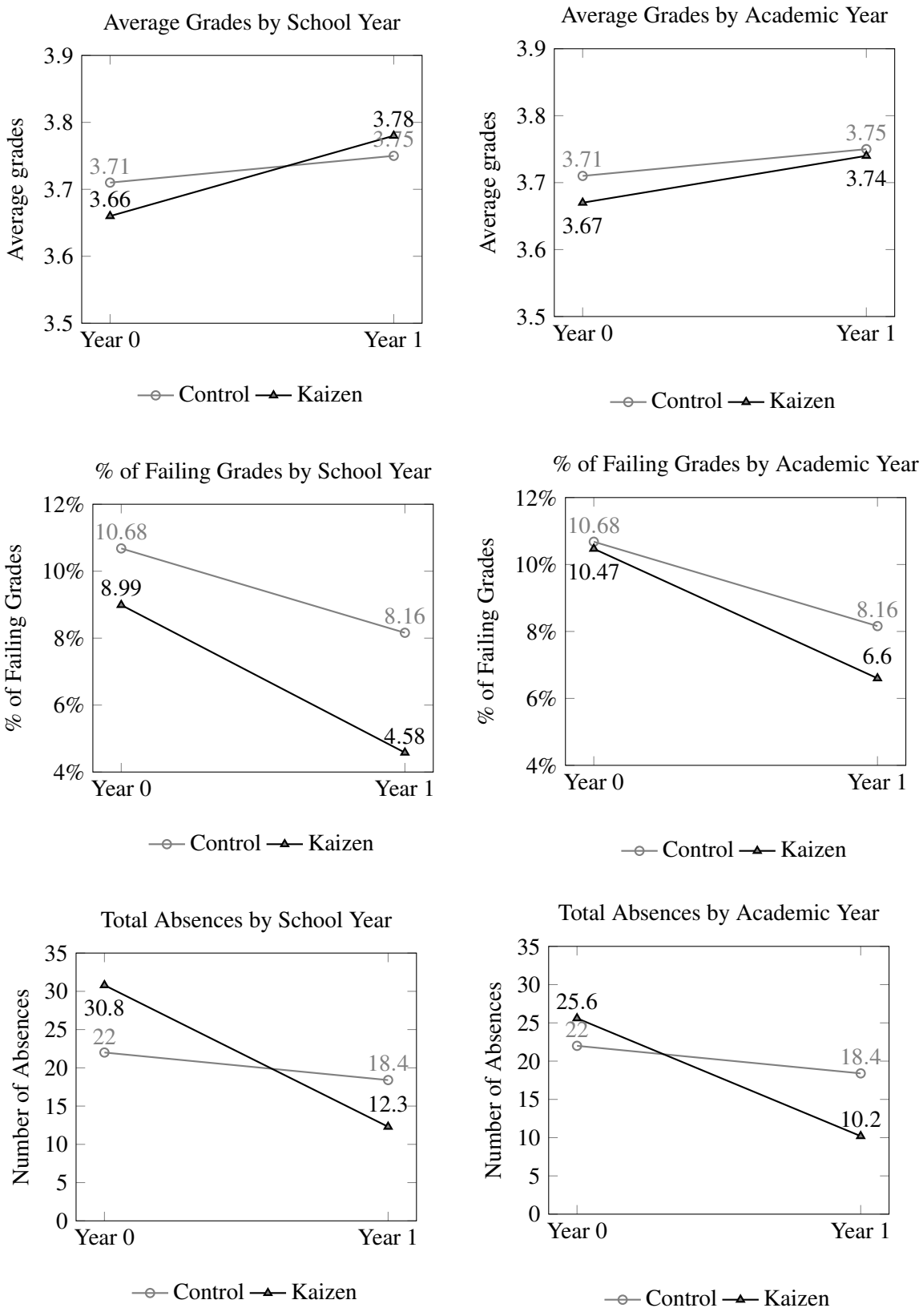


Figure 5.1: Evolution of Control and Kaizen for Average Grades, % of Failing Grades and Total Absences per student per year

Table 5.1: Kaizen vs Control Improvement

Indicator	**School Year**	**Academic Year**
Average Grades	69.70%	38.60%
% of Failing Grades	42.82%	34.64%
Total Absences	73.06%	73.14%

Two other significant external factors must be considered when interpreting the results of this analysis, as they may have influenced student performance independently of the program's effects. First, the aftermath of the COVID-19 pandemic continued to impact the Portuguese education system well into the 2021/2022 academic year.

As a result, some of the observed improvements during this period may be linked to grade inflation in the immediate post-pandemic context, where academic standards were relaxed and most students were not retained despite underperformance. This artificial elevation in grades may have distorted baseline values, potentially exaggerating perceived gains in the early years and complicating comparisons with subsequent periods, when grading practices and performance levels began to stabilize.

Second, the period under analysis was marked by political instability in Portugal, with frequent shifts in educational policy from year to year. These policy changes affected curriculum expectations, resource allocation, and the degree of autonomy available to school leadership. In practice, this created uncertainty and inconsistency in how policies were interpreted and implemented at the school level, often influencing the capacity of teachers and administrators to sustain long-term planning and effective program implementation. These systemic disruptions should be acknowledged as they may have introduced variability in program delivery or in the school environment, affecting student outcomes across different years of the project.

5.2 SROI

In today's resource-constrained environment, organizations operating in the social sector face increasing pressure to demonstrate not only that their programs work, but also that they are worth the investment. Social Return on Investment (SROI) has emerged as a powerful methodology for meeting this demand. Unlike traditional financial assessments, which often fail to capture the broader impacts of social programs, SROI translates a variety of social outcomes - such as improved education, health, or community well-being - into a common economic metric.

This method allows stakeholders to understand the broader societal value generated by an initiative, expressed as a ratio of benefits to investment. For instance, a program with an SROI of 5:1 is understood to deliver five dollars in social value for every dollar spent. Such analyses are increasingly used by foundations, governments, and community organizations to guide funding decisions and optimize program design.

The utility of SROI lies in its ability to make the intangible tangible. Social programs, particularly in education, health, and community development, frequently produce outcomes that are

difficult to quantify in financial terms. SROI addresses this challenge by assigning proxy monetary values to non-market outcomes, allowing organizations to evaluate, compare, and communicate the real-world effects of their initiatives in economic terms.

5.2.1 Kaizen Education SROI

Education is one of the most significant and socially valuable investments a society can make. Its long-term effects are well-documented: higher lifetime earnings, lower crime rates, better health outcomes, and increased civic participation. Yet, the benefits of educational initiatives are often underrepresented in cost-based evaluations. SROI provides a means of capturing these benefits holistically and presenting them in a way that resonates with decision-makers.

The three previously defined Key Performance Indicators (KPIs) were found to be particularly useful for the Social Return on Investment (SROI) analysis, as they provide measurable and meaningful evidence of the program's educational and social impact:

- Average grades: A strong indicator of academic progression, reflecting both student learning and instructional quality.
- Failure rates: Highlight students at risk and serve as a proxy for future remediation or intervention costs.
- Student absences: Correlate closely with student engagement and risk of dropping out, both of which carry significant long-term economic implications.

To conduct a meaningful SROI analysis, it is necessary to assign financial proxies to the educational outcomes being measured. These proxies are based on research linking educational performance to long-term socioeconomic benefits.

Used wisely, SROI is more than a reporting tool - it becomes a strategic asset for educational planning, advocacy, and sustainability. It provides evidence that educational investments yield substantial and measurable returns, reinforcing the argument that supporting students is not only ethically imperative but also economically sound.

This is calculated using the following formula:

$$\text{SROI} = \frac{\text{Total value of Social Benefits generated}}{\text{Total cost of investment}} \tag{5.1}$$

As previously mentioned, within the scope of the project implemented in primary education, a significant improvement in students' average grades was observed. This improvement is particularly relevant, not only as an indicator of immediate success, but also for its long-term effects on students' educational and economic trajectories.

According to the study by Avvisati et al. (2022), published in the *International Journal of Social Economics*, there is a direct link between academic performance in the early years of schooling

and the likelihood of progressing to higher levels of education. Students with higher grades in basic education are more likely to complete secondary school and pursue higher education, which significantly influences their lifetime earnings.

This phenomenon occurs because academic performance acts as a filter for educational progression: students with better results tend to face fewer barriers, are more motivated, and receive more support from families and teachers, which facilitates the transition to higher levels of education. To quantify this impact within the SROI analysis, the average salary differences associated with attained education levels were analyzed, as shown in Tables C.1 and C.2.

Based on this data, it can be concluded that each additional year of schooling represents, on average, an increase of €2,257 in annual income per individual. Therefore, by promoting the increase in average grades in primary education, the *Kaizen Education* project directly contributes to increasing the probability of completing higher education levels, resulting in a tangible future economic benefit for students.

This relationship between early academic performance, educational continuity, and future income constitutes a central element in justifying the social return on investment of the project. Improving grades is not just an academic gain: it is an investment with long-term multiplier effects - for students, their families, and society as a whole.

Regarding the second indicator, the reduction in the percentage of failing grades, this progress can also be linked to a direct social and economic impact, since grade repetition tends to occur early, often associated with failure in core subjects. This confirms the close relationship between the number of failing grades and the probability of retention. The author Alexander et al. (2003) analyzes the effects of grade retention in primary education and highlight that, although repetition is traditionally seen as a measure to improve academic performance, it can have long-lasting adverse effects, such as increasing the risk of school dropout and perpetuating educational inequalities.

Thus, by reducing the number of failing grades, the project acts preventively in reducing the number of retentions, which in turn generates direct savings for the education system. Each avoided retention corresponds to a reduction in the total cost of one school year per student, including human resources, materials, and social support.

This impact can be objectively incorporated into the SROI calculation, through the estimated number of retentions avoided and the average annual cost per student, reflecting both pedagogical benefits and direct economic gains from the intervention.

Finally, the third indicator - the reduction in the number of absences - is arguably the most directly measurable in economic terms. Each missed lesson represents an educational investment that is lost, both in terms of human resources (teacher time) and infrastructure and pedagogical planning.

By calculating the cost of a single lesson per student (Table C.6), it becomes possible to quantify the effective value associated with each avoided absence. This approach allows for a very objective assessment of the return on the intervention: fewer absences mean better use of invested

educational resources and, potentially, better academic outcomes and reduced risk of dropout in the medium to long term.

In this sense, by improving student attendance, the project not only ensures greater effectiveness of the teaching-learning process, but also improves the economic efficiency of the education system. This impact can be directly incorporated into the SROI calculation, considering:

- the number of class periods recovered due to improved attendance,
- multiplied by the estimated cost of each class period per student.

Thus, it becomes clear how a simple operational indicator - attendance - can be translated into concrete financial gains. All of this has been compiled in the tables in Appendix C and in the tables below.

Table 5.2: Proxy values used in the Kaizen Education SROI analysis

Outcome	**KPI**	**Proxy Value per Student**	**Justification**
Improved academic grades	Average grade increase in 83% of the cases	2,257€ per additional year of school	Correlated with increased graduation rates and future earnings
Reduced failure rates	Decrease of 32% in negative scores	6,458€ per repetition	Avoided cost of grade repetition
Reduced absenteeism	73% reduction in absences	5.98€ per absence	Linked to lower dropout rates and daily savings

Table 5.3: SROI Benefit per Component

Component	**Proxy Value per Student**	**Total Benefit**
Improved Grades	€2,257 per additional year of school	€339,679
Reduced Failures	€6,458 per repetition	€11,344
Reduced Absenteeism	€5.98 per absence	€509,267

Table 5.4: SROI Calculation Summary

Description	**Amount**
Total Estimated Value	€860,290
Investment	€70,300
SROI	12.24

It should be noted that the calculations below are made over a period of 4 years, meaning this is the average return of the project up to this point. Since the first year of implementation was

carried out exclusively during the participants' free time, without funding, its impact was also smaller (fewer resources).

The project produced an SROI ratio of 12.24, meaning that for every €1 invested, the program generated approximately €12.24 in social value. This is a remarkably high return, especially within the educational sector, and places the program well above typical public or social investment benchmarks.

The most significant monetized benefit was the reduction in absenteeism, which accounted for €509,267 of the total benefit - more than two-thirds of the overall social value generated. This suggests that the intervention was particularly effective in improving student engagement and attendance, which is often correlated with long-term academic success, reduced dropout risk, and better socio-emotional outcomes.

While the calculated benefit from reducing student failures in this SROI analysis appears relatively modest - totalling just €11,344 - this figure should be interpreted considering the sample's educational level. The analysis focused on students in primary education, a stage where failure and dropout rates are historically low in Portugal. National statistics report an average failure rate of approximately 1.8% in primary school (Table C.3) , meaning that relatively few students repeat a year, and consequently, the opportunity for immediate cost savings in this area is limited.

However, this dynamic is expected to shift significantly in later stages of schooling. In lower secondary and particularly in upper secondary, failure rates can rise to 10% or higher, especially in public schools with higher social vulnerability. Therefore, as the students currently engaged in the program progress through the education system, the long-term benefit of reducing grade repetition is likely to increase substantially. Future SROI analyses should take this into account, potentially incorporating projected values based on cohort progression.

Therefore, the results presented here could be viewed as conservative estimates of the program's value. Future evaluations that follow these same students into higher levels of education, or that adjust proxy values to reflect the higher risk profile of TEIP schools, are likely to reveal an even stronger return on investment.

Chapter 6

Next Steps

Building upon the promising results achieved thus far and the establishment of *Kaizen Education* as a transformative movement within school culture, the next phase will concentrate on broadening and deepening the evaluation of its impact. It is essential to advance towards a more structured and systematic approach to data collection, one that captures not only quantitative indicators - such as improvements in student grades and reductions in failure rates - but also qualitative outcomes related to the engagement of educational communities and the cultural shifts occurring within schools.

Moreover, it will be crucial to strengthen collaboration with schools, local authorities, and institutional partners to ensure access to reliable, disaggregated, and comparable datasets. This will facilitate more precise and continuous analysis, addressing the challenges previously identified in data collection and organization. Ultimately, this approach aims to establish a sustainable evaluation system that is closely aligned with the actual needs and realities of the schools involved.

As part of the next steps, it will be important to monitor the project's steady growth over time. This includes tracking metrics such as the increase in the number of students reached, as well as the number and geographic distribution of participating schools. These indicators will provide valuable insights into the project's expansion and help guide future strategic planning and resource allocation. Furthermore, additional aspects related to the assessment of *Kaizen*'s impact in education will be explored in the upcoming subchapters.

6.1 Quantitative vs Qualitative Indicators

The current approach to evaluating educational impact in Portugal has made effective use of quantitative indicators - such as average grades, failure rates, and absenteeism - which are essential for tracking academic progress and conducting structured analyses like the SROI. These metrics have provided a solid basis for understanding the program's effectiveness and communicating its value in economic terms.

At the same time, there is a clear opportunity to build upon these efforts by further developing the qualitative dimensions of evaluation, particularly in areas that are more difficult to capture

through numbers alone, such as student motivation, classroom climate, and socio-emotional development.

Importantly, this opportunity does not require starting from scratch. Surveys are already being administered within the context of the *Kaizen Education Movement*, offering valuable insight into stakeholder perceptions and student experiences. These existing tools can now be consolidated, refined, and expanded to include a broader and more diverse group of respondents - such as students across different grade levels, teachers, school leaders, and even families. Doing so would strengthen the representativeness of the data and help surface trends that may not be visible through academic indicators alone.

In particular, future iterations of these surveys could aim to:

- Track changes over time in attitudes toward learning, feelings of belonging, and perceived support from teachers;
- Differentiate responses by demographic or contextual variables, such as grade level, gender, or school location, to identify specific needs or barriers;
- Inform the design of targeted interventions, using feedback from those most directly impacted to adapt strategies in real time.

By combining these qualitative insights with the robust quantitative results already being collected, the program's evaluation framework would move closer to a comprehensive, mixed-methods model. This would not only enhance the depth of analysis but also align with the evolving goals of the intervention - particularly in TEIP schools, where academic progress is often tightly linked with social and emotional factors.

6.2 The Difference-in-Differences (DiD) Statistical approach

As the *Kaizen Education Movement* continues to consolidate its presence within Portugal's educational landscape, it is entering a new phase that demands greater rigor, recognition, and replicability. What began as an ambitious and holistic educational intervention is now transitioning from implementation to institutionalization. This shift brings with it the need for more advanced evaluation tools, not as an academic exercise, but as a strategic necessity for the movement's future.

This is not merely a suggestion for further research; it is part of an active plan already in motion. Over the past academic year, the need for a more robust statistical framework has been a recurring theme in discussions with education specialists, in working sessions with school leadership, and during formal presentations at national congresses and sectoral meetings. These conversations have consistently pointed toward one conclusion: for *Kaizen Education* to establish itself not only as a successful intervention, but as a reference model in Portugal and beyond, it must adopt and demonstrate evidence-based practices grounded in statistical rigor.

To this end, implementing quasi-experimental impact evaluation methods is the logical next step. Among these, the Difference-in-Differences (DiD) approach stands out as particularly well-suited to the structure, scale, and timeline of the *Kaizen Education* initiative. DiD offers a means to statistically confirm the hypotheses that have emerged from the movement's practice over the past years: that improved student performance, reduced absenteeism, and higher levels of engagement are causally linked to the pedagogical and structural innovations introduced by the program.

By adopting DiD, *Kaizen Education* not only meets growing expectations for transparency and measurable impact but also positions itself at the forefront of educational innovation in Europe. In a time when policy decisions increasingly rely on demonstrable outcomes, and when funders and institutions require evidence of scalable success, using DiD signals a commitment to both scientific integrity and strategic ambition. It reflects a movement that is not content with anecdotal success, but that aspires to lead through verified results and continuous improvement.

What follows in this chapter is a detailed explanation of the Difference-in-Differences method, its suitability for this educational context, and its potential to shape the next stage of *Kaizen Education*'s evolution - transforming it from a promising intervention into a national and international model of educational transformation.

Assessing the causal impact of educational interventions in real-world contexts poses a methodological challenge, particularly where randomization is not feasible due to ethical, logistical, or institutional constraints. In this context, the Difference-in-Differences (DiD) approach has become a prominent quasi-experimental technique, allowing researchers to draw meaningful causal inferences from observational data. As noted by Gertler et al. (2016), DiD "cleans" the before-and-after differences of time-varying confounders by subtracting similar trends in the comparison group, thereby offering a more accurate estimate of the intervention's true effect.

DiD compares changes in outcomes over time between a treatment group - which is exposed to an intervention - and a comparison group - which is not. The technique estimates what would have happened to the treatment group in the absence of the intervention by assuming that the outcome trend in the comparison group reflects the counterfactual. In doing so, DiD corrects for selection bias due to time-invariant differences between groups and removes the influence of broader, time-varying factors by differencing changes in both groups. This combination of a before–after comparison with a control–treatment distinction significantly improves on either method used in isolation.

Such methods have already proven valuable in diverse contexts. Moura et al. (2022) used DiD to evaluate the impact of financial regulation in the Eurozone, accounting for structural efficiency trends in banking. Burzacchi et al. (2024) employed the approach to estimate the effect of commuting time on academic performance, capturing changes across treated and untreated groups of students at Politecnico di Milano. Likewise, in education policy, Sneyers and De Witte (2018) emphasized the methodological strength of DiD among quasi-experimental designs in their meta-analysis of school-based interventions. Considering the widespread grade inflation identified by François and De Witte (2025) in the post-COVID era, the DiD model's ability to filter out system-wide shocks becomes even more crucial in ensuring the validity of outcome-based evaluations.

In the context of the *Kaizen Education*, DiD offers an especially relevant analytical framework. The Movement's phased implementation across school years and student cohorts creates natural conditions for comparison. By using prior-year cohorts or non-participating schools as controls, and by measuring outcomes such as grades, failure rates, and absenteeism before and after the intervention, DiD can provide robust causal estimates. These estimates remain valid even in the presence of national reforms, grading changes, or broader social developments, as long as those changes affect both groups equally - thus fulfilling the parallel trends assumption. Future research could, for instance, explore whether the program has differential impacts by gender, socioeconomic background, or initial academic standing, offering critical insights for policy targeting and resource allocation.

In practice, the *Kaizen Education* initiative already compares outcomes between *Kaizen* classes and matched control groups across successive cohorts. What is still lacking, however, is the formal statistical confirmation that a Difference-in-Differences framework can provide, along with the specialised technical expertise required to implement it rigorously. Embedding DiD analysis - and the requisite training or external support - into KE's evaluation cycle will turn its existing comparison design into evidence that meets the highest standards of causal inference, thereby converting promising descriptive results into verified impact estimates and solidifying the movement's credibility for stakeholders, funders, and policy-makers alike.

Chapter 7

Conclusion

Portugal has an extensive network of more than 4,500 public basic and secondary schools, which play a fundamental role in shaping new generations. However, this network faces persistent challenges that jeopardize academic success and equitable opportunities. Difficulties range from regular absenteeism, reported by half of the teachers, according to recent data (Diário de Notícias, 2025); to the progressive disengagement of students, often exacerbated by unfavourable socioeconomic contexts. These problems not only compromise students' academic trajectories, but also limit their personal and professional development throughout life.

It is in this context that projects like *Kaizen Education* prove truly transformative. With an approach centered on continuous improvement, valuing the potential of each student, and empowering educational teams, *Kaizen* does not act as an isolated solution but as an integrated response to structural gaps in the education system. Its goal is clear: to be present in every school across the country and to reach all students, regardless of their background or context. More than an ambition, this is a mission that guides every action, intervention plan, and evaluation cycle.

The results obtained so far are highly promising. Indicators such as improved grades, reduced failure rates, and increased school engagement demonstrate that it is possible, with strategy and consistency, to transform complex educational realities. But beyond the numbers, it is the cultural shift in schools - marked by greater involvement, organization, purpose, and collective meaning - that confirms the positive impact of the *Kaizen* approach.

By integrating planning practices, structured problem solving, error reduction, and a focus on common goals, the *Kaizen Education Movement* promotes skills that are essential not only throughout the school journey but also in adult life. These are the foundations of a more productive and fair society, better prepared for future challenges.

This dissertation thus has sought to demonstrate how it is possible to combine pedagogical vision, social responsibility, and organizational discipline to renew public schools from within. *Kaizen Education* does not aim to replace the system but to complement and strengthen it, progressively eliminating its shortcomings and enhancing its resources. With a firm commitment to continuous improvement, this movement is building a new narrative for education in Portugal - a narrative where everyone counts, where every school matters, and where no student is left behind.

It is also important to highlight the significant challenges faced throughout this research, especially in the collection and organization of detailed and consistent databases. Obtaining reliable, disaggregated, and comparable data remains one of the greatest obstacles to developing impact studies in educational contexts. Accessing this information required ongoing coordination with schools, local entities, and institutional partners, which only reinforces the importance of creating more open, interoperable, and continuously assessment-focused systems.

Furthermore, this work was developed in a context still marked by institutional uncertainties and resource limitations, which makes the results and structure of the research even more relevant. The journey was demanding but allowed overcoming barriers that often hinder progress in applied educational science.

In a national landscape where projects rigorously evaluating their impact in school settings are rare, *Kaizen Education* stands out as an innovative proposal, with ambition for scale and a sustained, measurable improvement approach. This dissertation is therefore not only an impact analysis but also a firm step toward the strengthening and enhancement of public education in Portugal.

Bibliography

Alexander, K. L., Entwisle, D. R., and Dauber, S. L. (2003). *On the Success of Failure: A Reassessment of the Effects of Retention in the Primary School Grades*. Cambridge University Press, 2nd edition.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association, Washington, DC.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Ariel, B., Sutherland, A., and Bland, M. (2021). The trick does not work if you have already seen the gorilla: how anticipatory effects contaminate pre-treatment measures in field experiments. *Journal of Experimental Criminology*, 17:55–66.

Avvisati, F., Bertoletti, A., and Charbonnier, E. (2022). How do student performance and school climate shape long-term education outcomes? *International Journal of Social Economics*, 49(11):1594–1614.

Bititci, U. S., Garengo, P., Ates, A., and Nudurupati, S. S. (2015). Value of maturity models in performance measurement systems. *International Journal of Production Research*, 53(10):3062–3085.

Brookhart, S. M. (2013). *How to Create and Use Rubrics for Formative Assessment and Grading*. ASCD.

Bryk, A. S., Gomez, L. M., Grunow, A., and LeMahieu, P. G. (2015). *Learning to Improve: How America's Schools Can Get Better at Getting Better*. Harvard Education Press. pp. 12–17.

Burzacchi, A., Rossi, L., Agasisti, T., Paganoni, A. M., and Vantini, S. (2024). Urban mobility and learning: Analyzing the influence of commuting time on students' gpa at politecnico di milano. *Studies in Higher Education*. Advance online publication.

Cook, T. D. and Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Houghton Mifflin. Ch. 1 pp. 37–40; Ch. 3 pp. 80–85.

Cousins, J. B. and Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, 1998(80):5–23.

Cronbach, L. J. (1971). Test validation. In Thorndike, R. L., editor, *Educational Measurement*, pages 443–507. American Council on Education, 2nd edition.

de Waal, A. A. (2003). Behavioral factors important for the successful implementation and use of performance management systems. *Management Decision*, 41(8):688–697.

Deming, W. E. (1986). *Out of the Crisis*. Massachusetts Institute of Technology, Center for Advanced Engineering Study.

DGEEC (2023). Caracterização geral - alunos matriculados/inscritos.

Diário de Notícias (2025). Metade dos professores tem alunos que faltam frequentemente às aulas. *Diário de Notícias*.

Durlak, J. A. and DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3-4):327–350.

Earl, L. and Katz, S. (2002). Leading schools in a data-rich world. In Leithwood, K. and Hallinger, P., editors, *Second International Handbook of Educational Leadership and Administration*, volume 8 of *Springer International Handbooks of Education*, pages 645–650. Springer.

Education, K. (2025). Internal pilot evaluation report: Impact indicators and implementation data. Unpublished internal document.

EduStat (2022). Ganhos mensais por nível de escolaridade – Indicador 105.

EduStat (2022). Infostat: Indicadores-chave por agrupamento de escolas. `https://www.edustat.pt/Infostat/InfostatDetail?ID=14`.

Edustat (2023). Indicator 401.

EduStat (2025). Despesa da administração pública em educação.

Escola do Falcão (2022). Implementação kaizen na escola do falcão (2021/2022). Internal document.

Fernandes, D. (2009). Educational assessment in portugal. *Assessment in Education: Principles, Policy Practice*, 16(2):241–261.

Franco-Santos, M., Kennerley, M., Micheli, P., Martinez, V., Mason, S., Marr, B., Gray, D., and Neely, A. (2007). Towards a definition of a business performance measurement system. *International Journal of Operations & Production Management*, 27(8):784–801.

François, M. and De Witte, K. (2025). A decade of grade inflation boosted by the covid-19 pandemic—an empirical analysis of a top european university. *British Educational Research Journal*.

Fullan, M. (2016). *The New Meaning of Educational Change*. Teachers College Press, 5th edition.

Garet, M. S., Mitrano, S., Eisner, R., Kochanek, J. R., Jones, K. T., Ibis, M., and Estrada, S. (2021). Continuous improvement in education settings: A literature review. American Institutes for Research. `https://www.air.org/sites/default/files/2022-03/Continuous-Improvement-in-Education-Settings-A-Literature-Review-August-2021.pdf`.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., and Vermeersch, C. M. J. (2016). *Impact Evaluation in Practice*. World Bank, second edition. Chapter 5, pp. 155–160.

Ghalayini, A. M. and Noble, J. S. (1996). The changing basis of performance measurement. *International Journal of Operations & Production Management*, 16(8):63–80.

Ghorna, L., Elly, A., Oladele, S., and Ahsun, A. (2024). The importance of data quality in data-driven decision-making. *Research Gate*, pages 2–3.

Görgens, M. and Kusek, J. Z. (2009). Making monitoring and evaluation systems work: A capacity development toolkit. *World Bank Publications*, pages 90–95.

Graham, L. J., McCarthy, T., Killingly, C., Tancredi, H., and Poed, S. (2022). Front. psychol. *Frontiers in Psychology*, 13:5–6.

Hanushek, E. A. (2005). Why quality matters in education. *Finance & Development*, 42(2).

Imai, M. (2012). *Gemba Kaizen: A Commonsense Approach to a Continuous Improvement Strategy*. McGraw Hill Professional, second edition.

Instituto Nacional de Estatística (2023). Indicadores de educação.

Kane, T. J. and Staiger, D. O. (2012). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. Research paper, MET Project, Bill Melinda Gates Foundation. pp. 10–20.

Kaplan, R. S. and Norton, D. P. (1996). *The Balanced Scorecard: Translating Strategy into Action*. Harvard Business Press.

Langley, G. J., Moen, R. D., Nolan, K. M., Nolan, T. W., Norman, C. L., and Provost, L. P. (2009). *The Improvement Guide: A Practical Approach to Enhancing Organizational Performance*. Jossey-Bass, 2nd edition.

Leite, A., Kislaya, I., Machado, A., Aguiar, P., Nunes, B., and Dias, C. M. (2023). Use of quasi-experimental studies to evaluate causal effects of public health interventions in portugal: A scoping review. *Population Medicine*, 5(Supplement):A1862.

Martinez, L. and Hayes, C. D. (2013). Measuring social return on investment for community schools: A case study. *The Finance Project & The Children's Aid Society*, pages 3–6. `https://files.eric.ed.gov/fulltext/ED561996.pdf`.

Mayne, J. (2008). Building an evaluative culture for effective evaluation and results management. *Canadian Journal of Program Evaluation*, 24(2):137–150.

Mayne, J. (2010). Building an evaluative culture: The key to effective evaluation and results management. *Canadian Journal of Program Evaluation*, 24(2):1–30.

Ministério da Educação (2023). Organização do sistema educativo português. `https://www.portugal.gov.pt`.

Moura, P., Barbosa, F., Alves, C., and Camanho, A. S. (2022). The impact of the single supervisory mechanism on eurozone banking: The assessment of trends in efficiency and frontier position. *Journal of Banking Regulation*, 23(3):165–185.

Neely, A., Adams, C., and Kennerley, M. (2002). *The Performance Prism: The Scorecard for Measuring and Managing Business Success*. Financial Times/Prentice Hall.

OECD (2010). The high cost of low educational performance. https://www.oecd.org/en/publications/the-high-cost-of-low-educational-performance_9789264077485-en.html, see pp. 10–20.

OECD (2013). *Evaluation and Assessment Frameworks for Improving School Outcomes: Policy Review*. OECD Publishing.

OECD (2019a). *Balancing School Choice and Equity: An International Perspective Based on PISA*. PISA. OECD Publishing, Paris.

OECD (2019b). *Better Criteria for Better Evaluation: Revised Evaluation Criteria Definitions and Principles for Use*. OECD Publishing, Paris. pp. 3–7).

OECD (2020). *TALIS 2018 Results (Volume II): Teachers and School Leaders as Valued Professionals*. OECD Publishing, Paris.

OECD (2021). *Applying Evaluation Criteria Thoughtfully*. OECD Publishing, Paris.

OECD (2022a). How to select indicators that support the implementation of education policies. Technical Report No. 1, OECD Publishing, Paris.

OECD (2022b). *Review of Inclusive Education in Portugal*. Reviews of National Policies for Education. OECD Publishing, Paris.

OECD (2024). Education at a glance 2024: Key indicators.

OECD (2024). *Education at a Glance 2024: OECD Indicators*. OECD Publishing, Paris.

Park, S., Hironaka, S., Carver, P., and Nordstrum, L. (2013). Continuous improvement in education. White paper, Carnegie Foundation for the Advancement of Teaching. pp. 1, 3–5.

Parmenter, D. (2015). *Key Performance Indicators: Developing, Implementing, and Using Winning KPIs*. Wiley, 3rd edition.

Patton, M. Q. (2002). *Qualitative Research & Evaluation Methods*. Sage Publications, Thousand Oaks, CA, 3rd edition. pp. 340–360.

Pawson, R. (2003). Nothing as practical as a good theory. *Evaluation*, 9(4):471–490.

Preskill, H. and Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29(4):443–459.

Psacharopoulos, G. and Patrinos, H. A. (2018). Returns to investment in education: A decennial review of the global literature. *Education Economics*, 26(5):445–458.

Reardon, S. F. (2019). Educational opportunity in early and middle childhood: Using full population administrative data to study variation by place and age. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 5(2):40–68.

Rossi, P. H., Lipsey, M. W., and Freeman, H. E. (2004). *Evaluation: A Systematic Approach*. Sage Publications, Thousand Oaks, CA, 7th edition.

Sakız, H., Özdaş, F., Göksu, , and Ekinci, A. (2021). A longitudinal analysis of academic achievement and its correlates in higher education. *SAGE Open*, 11(4).

Santiago, P., Donaldson, G., Looney, A., and Nusche, D. (2012). *OECD Reviews of Evaluation and Assessment in Education: Portugal 2012*. OECD Publishing. pp. 10–15.

Scriven, M. (1991). *Evaluation Thesaurus*. Sage Publications, Inc, 4th edition.

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin. pp. 15–17.

Sneyers, E. and De Witte, K. (2018). The effectiveness of school-based interventions in secondary education: A meta-analysis of 30 years of research. *Applied Economics Perspectives and Policy*, 40(3):478–497.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684):677–680.

Taticchi, P., Tonelli, F., and Pasqualino, R. (2013). Performance measurement of sustainable supply chains: A literature review and a research agenda. *International Journal of Productivity and Performance Management*, 62(8):782–804.

UNESCO (2015). *Rethinking Education: Towards a Global Common Good?* United Nations Educational, Scientific and Cultural Organization.

Van den Broeck, J., Cunningham, S. A., Eeckels, R., and Herbst, K. (2005). Data cleaning: Detecting, diagnosing, and editing data abnormalities. *PLOS Medicine*, 2(10).

VER (2023). Dinheiro não compra educação.

Vincent-Lancrin, S. et al. (2019). Measuring innovation in education 2019: What has changed in the classroom?

White, H. and Sabarwal, S. (2014). Quasi-experimental design and methods. Methodological briefs: Impact evaluation, 8, UNICEF Office of Research – Innocenti. pp. 1–6.

Appendix A

Auxiliary calculations of the model's KPIs

Table A.1: Summary of average grades by year and group (Control vs Kaizen)

Year 0	**Year 1**	**Control Average Grades**			**Kaizen Average Grades**		
		Year 0	Year 1	Change	Year 0	Year 1	Change
1st year	2nd year	3.83	3.82	-0.2%	3.73	3.78	1.4%
2nd year	3rd year	3.65	3.69	1.2%	3.66	3.67	0.2%
3rd year	4th year	3.64	3.72	1.9%	3.58	3.90	8.0%
Total School Year		3.71	3.75	1.0%	3.66	3.78	3.2%
2021/2022		3.67	3.68	0.3%	3.51	3.71	5.0%
2022/2023		3.64	3.67	0.7%	3.85	3.71	-3.5%
2023/2024		3.82	3.89	1.9%	3.66	3.79	3.4%
Total Academic Year		3.71	3.75	1.0%	3.67	3.74	1.6%
2024/2025 *		3.68	3.68	0.1%	3.50	3.20	-7.4%

Table A.2: Summary of failing grades by year and group (Control vs Kaizen)

Year 0	**Year 1**	**Control Failing Grades**			**Kaizen Failing Grades**		
		Year 0	Year 1	Change	Year 0	Year 1	Change
1st year	2nd year	11.05%	8.38%	-2.7%	10.09%	9.74%	-0.4%
2nd year	3rd year	11.06%	9.01%	-2.1%	14.81%	4.00%	-10.8%
3rd year	4th year	9.93%	7.08%	-2.9%	2.08%	0.00%	-2.1%
Total School Year		10.68%	8.16%	-2.5%	8.99%	4.58%	-4.4%
2021/2022		9.83%	7.21%	-2.6%	14.81%	5.56%	-9.3%
2022/2023		10.71%	8.91%	-1.8%	10.50%	10.42%	-0.1%
2023/2024		11.50%	8.35%	-3.2%	6.09%	3.83%	-2.3%
Total Academic Year		10.68%	8.16%	-2.5%	10.47%	6.60%	-3.9%
2024/2025		7.93%	12.87%	4.9%	16.35%	15.66%	-0.7%

Table A.3: Summary of absences by year and group (Control vs Kaizen)

Year 0	**Year 1**	**Control Absences**			**Kaizen Absences**		
		Year 0	Year 1	Change	Year 0	Year 1	Change
1st year	2nd year	26.7	34.0	27.5%	26.0	17.5	-32.7%
2nd year	3rd year	21.3	12.3	-42.2%	13.5	7.5	-44.4%
3rd year	4th year	18.0	9.0	-50.0%	53.0	12.0	-77.4%
Total School Year		22.00	18.44	-16.2%	30.83	12.33	-60.0%
2021/2022		28.7	13.7	-52.3%	24.0	14.0	-41.7%
2022/2023		23.7	9.7	-59.2%	1.0	1.0	0.0%
2023/2024		13.7	32.0	134.1%	27.3	15.7	-42.7%
Total Academic Year		22.00	18.44	-16.2%	25.67	10.22	-60.2%
2024/2025		40.7	17.7	-56.6%	20.0	3.0	-85.0%

Table A.4: Total Kaizen Absences per student per Academic Year

Grade	2021/2022	2022/2023	2023/2024	2024/2025
2nd Year	0	99	135	158
3rd Year	63	0	106	125
4th Year	0	0	288	80
Total	63	99	529	363

Table A.5: Total Kaizen Absences Reduced (73%) per student per Academic Year

Grade	2021/2022	2022/2023	2023/2024	2024/2025
2nd Year	0	72	99	115
3rd Year	46	0	77	91
4th Year	0	0	210	58
Total	46	72	387	265

Table A.6: Total Kaizen Students per Academic Year

Grade	2021/2022	2022/2023	2023/2024	2024/2025
2nd Year	0	16	36	36
3rd Year	36	0	14	29
4th Year	0	0	16	32
Total	36	16	66	97

Table A.7: Total Saving per Academic Year (€)

Grade	2021/2022	2022/2023	2023/2024	2024/2025
2nd Year	0	6921	21234	244852
3rd Year	9909	0	6484	15838
4th Year	0	0	20133	11185
Total	9909	6921	47852	51876

Appendix B

Dashboard views

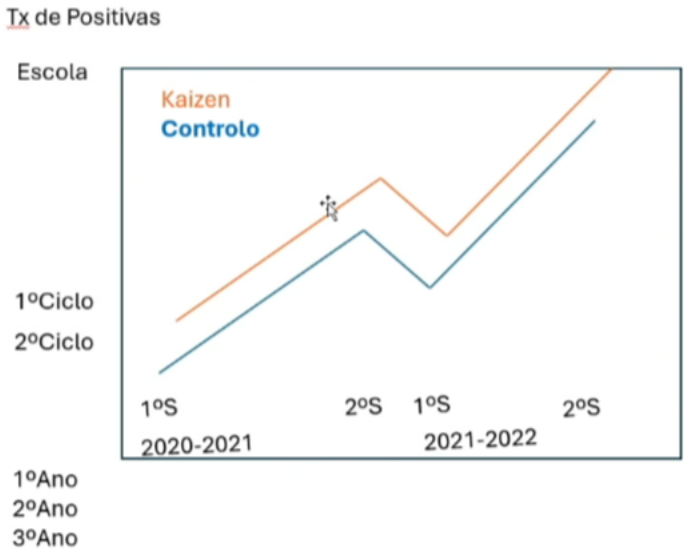


Figure B.1: Initial Dashboard Mock-up - Annual View

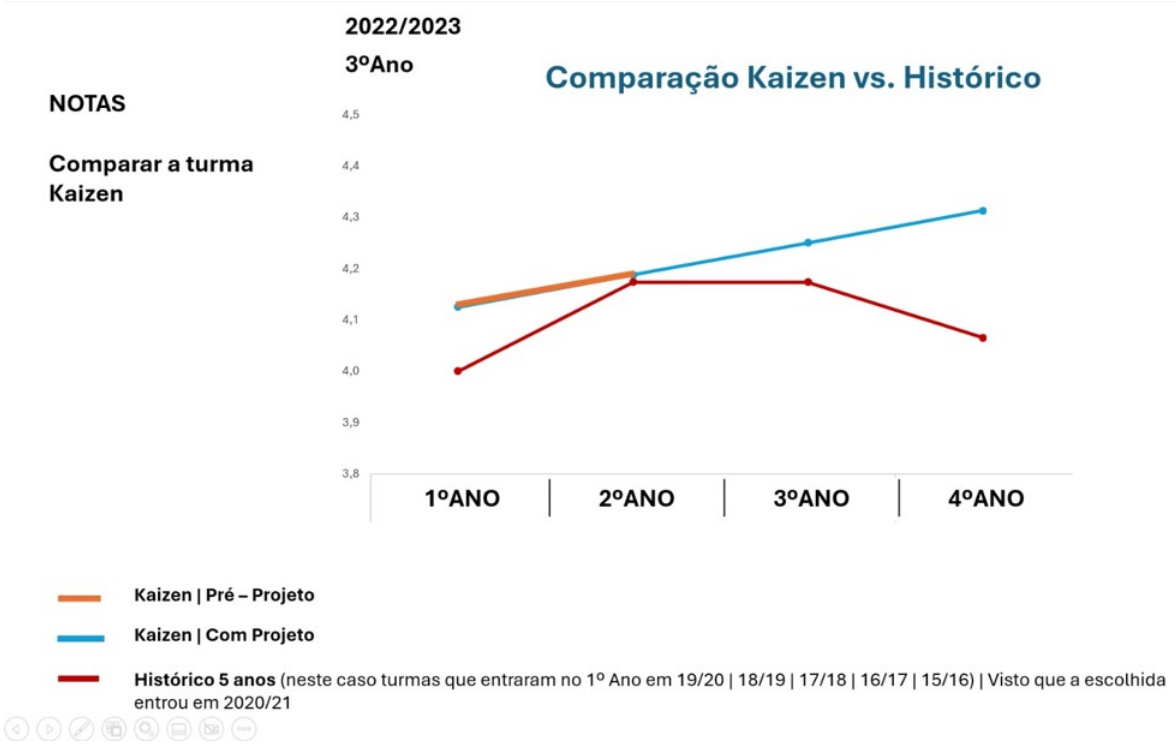


Figure B.2: Initial Dashboard Mock-up - Longitudinal View

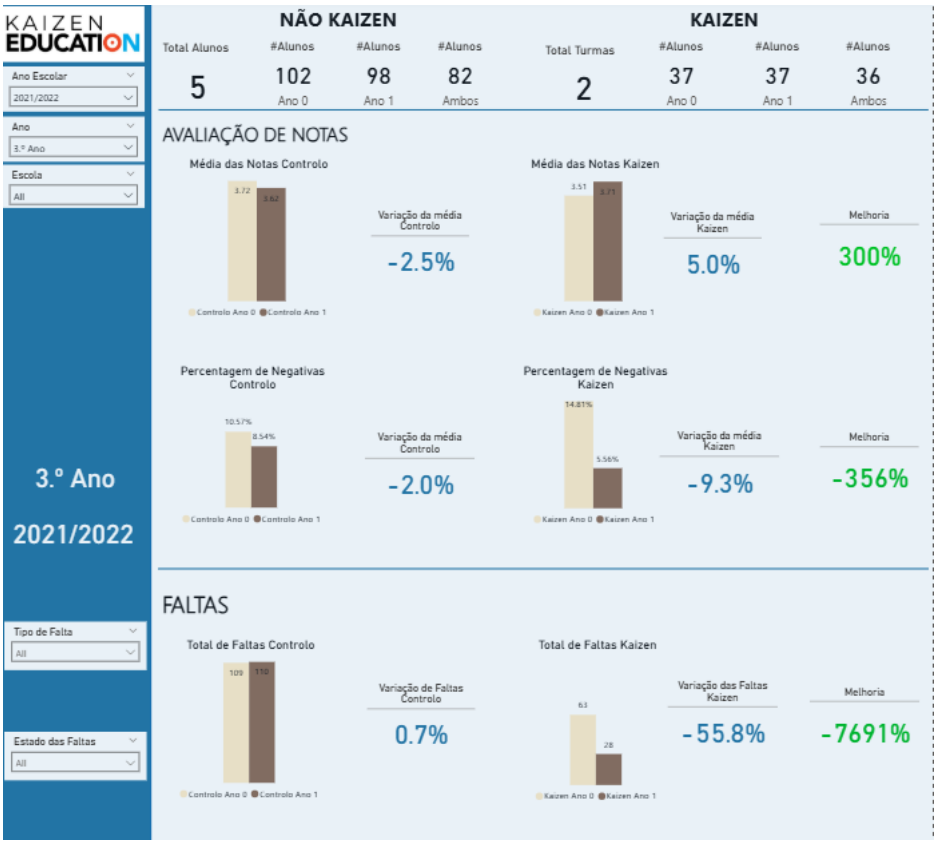


Figure B.3: Evolution and comparison of Kaizen vs Control from 2021/2022 to 2023/2024

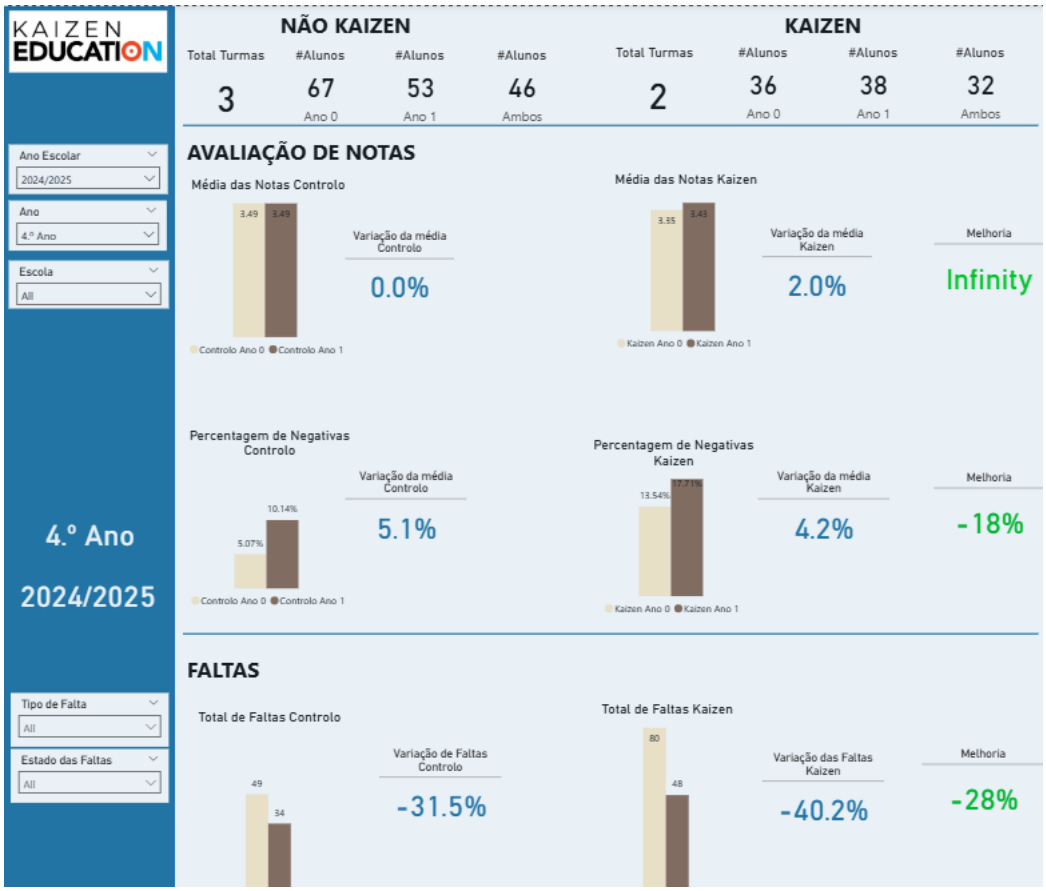


Figure B.4: Evolution and comparison of Kaizen vs Control in 2024/2025 (1st semester)

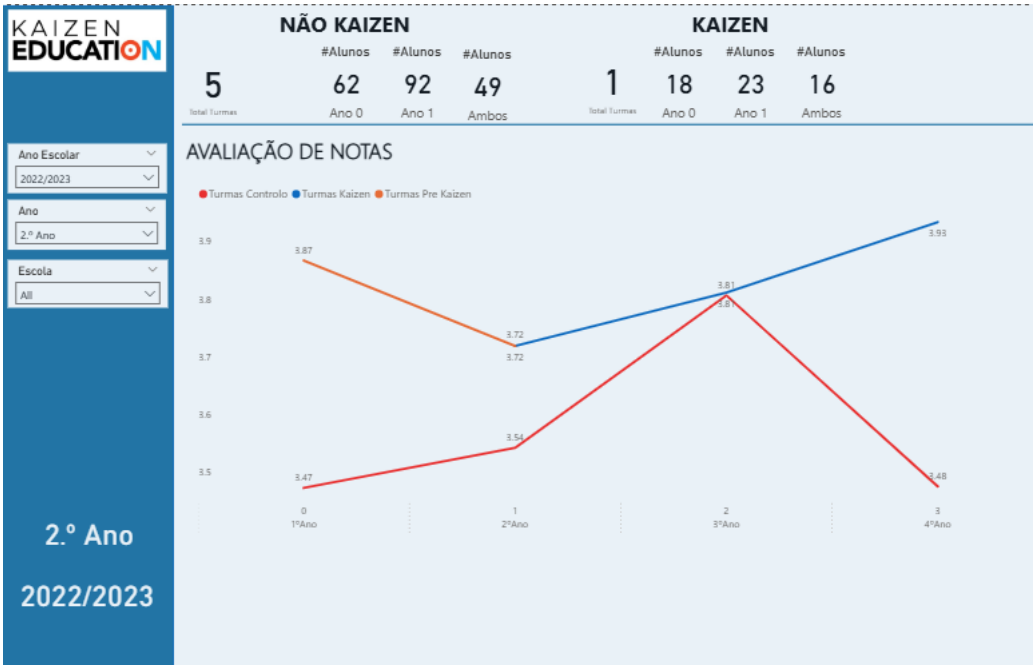


Figure B.5: Longitudinal overview of Kaizen vs 5 years past

Appendix C

SROI Calculations

Table C.1: Average annual salary increase by level of education

School Transition (years)	Salary Difference (€)	Difference (Years)	Increase per Year (€)
Basic (9) → Secondary (12)	2,045	3	682
Secondary (12) → Higher Education (15)	11,494	3	3,831

Table C.2: Weighted average of salary increase per additional year of schooling

Weighted average increase (€/year)	2,257

Table C.3: Retention costs and total cost by level of education (2022)

Level of Education	Retention Rate (%)	Total Students	Total Cost (€)
Primary (1st cycle)	1.8	388,316	7,038,616
Lower Secondary (2nd cycle)	3.1	210,345	6,566,340
Upper Secondary (3rd cycle)	6.2	346,788	21,651,361
High School (Secondary)* **	9.8	394,964	38,977,417

** Student retention in high school: Scientific-Humanities (9.3%) vs Professional (10.6%).*

*** Percentage of students in private basic education: approx. 13%; in private secondary: approx. 25%.*

[1]

[1]The data and analysis presented in this Appendix are based on information from the DGEEC (2023), the article "Dinheiro não compra educação" VER (2023), statistics from the Instituto Nacional de Estatística (2023), and indicators from Edustat (2023) and EduStat (2022)

Table C.4: Retention cost in the public sector by educational cycle (2022)

Level of Education	Retention Rate (%)	Total Students	Average Cost p/ Retention (€)	Total Cost (€)
Primary (1st cycle)	1.8	335,674	6,798	41,074,413
Lower Secondary (2nd cycle)	3.1	184,445	6,798	38,869,570
Upper Secondary (3rd cycle)	6.2	303,363	8,552	160,850,343
High School (Secondary)	9.8	296,371	9,064	263,258,060

Table C.5: Summary of retention costs and budget ratio

Indicator	Value
Government Budget for Education*	10,813,000,000€
Total retention cost	503,000,000 €
Ratio to Education Budget	4.7%

** Values obtained with EduStat (2025)*

Table C.6: Key Teaching Cost Indicators (2022)

Indicator	Value
Average cost per student (2022)	€6,458
Weekly teaching hours	25 hours
Minutes per week	1,500 minutes
Teaching periods per week	30 periods
Effective weeks per year	36 weeks
Teaching periods per year (per class)	1,080 periods
Cost per teaching period	€5.98

Appendix D

NPS Results

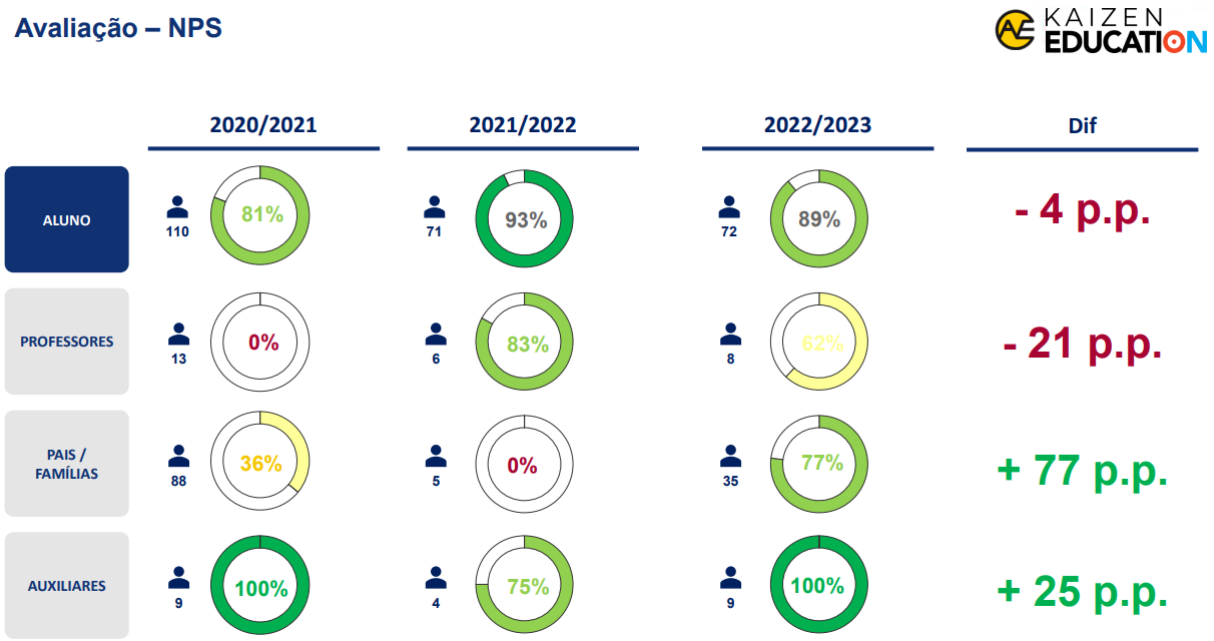


Figure D.1: Net Promoter Score (NPS) Results at Escola Básica do Falcão (2020/2021 to 2022/2023)

KAIZEN EDUCATION

	1. Estás contente com o que podes decidir na escola?	2. Estás contente com a ajuda dos teus professores?	3. Estás contente com os espaços da escola para trabalhares e brincares?	4. Estás contente com o material da escola para trabalhares e brincares?	5. Gostas da comida / espaço da cantina?	6. Estás contente com a quantidade de T.P.C.?	7. Como correm os teus recreios?	8. Sentes-te seguro na Escola?	9. Estás contente com os teus colegas?	10. Confias nos teus professores?	11. Gostas da importância que, em casa, te dão quando trazes boas notas/trabalhos?	12. Estás contente com a escola?
Total 22/23	74%	92%	79%	75%	29%	49%	64%	74%	74%	90%	86%	89%
Total 21/22	74%	97%	90%	87%	32%	49%	49%	75%	86%	96%	89%	94%
Total 20/21	74%	93%	75%	80%	38%	55%	48%	68%	74%	95%	92%	81%

22/23												
1º	53%	87%	53%	60%	29%	13%	33%	40%	47%	87%	71%	57%
2º	100%	100%	100%	100%	80%	80%	90%	90%	100%	100%	100%	100%
3º	63%	88%	87%	75%	20%	69%	63%	69%	63%	100%	88%	94%
4º	81%	94%	81%	74%	17%	45%	71%	87%	84%	84%	87%	97%

Figure D.2: Net Promoter Score (NPS) Results at Escola Básica do Falcão (2020/2021 to 2022/2023) - Detailed View

Appendix E

Examples of Kaizen Practices in Action

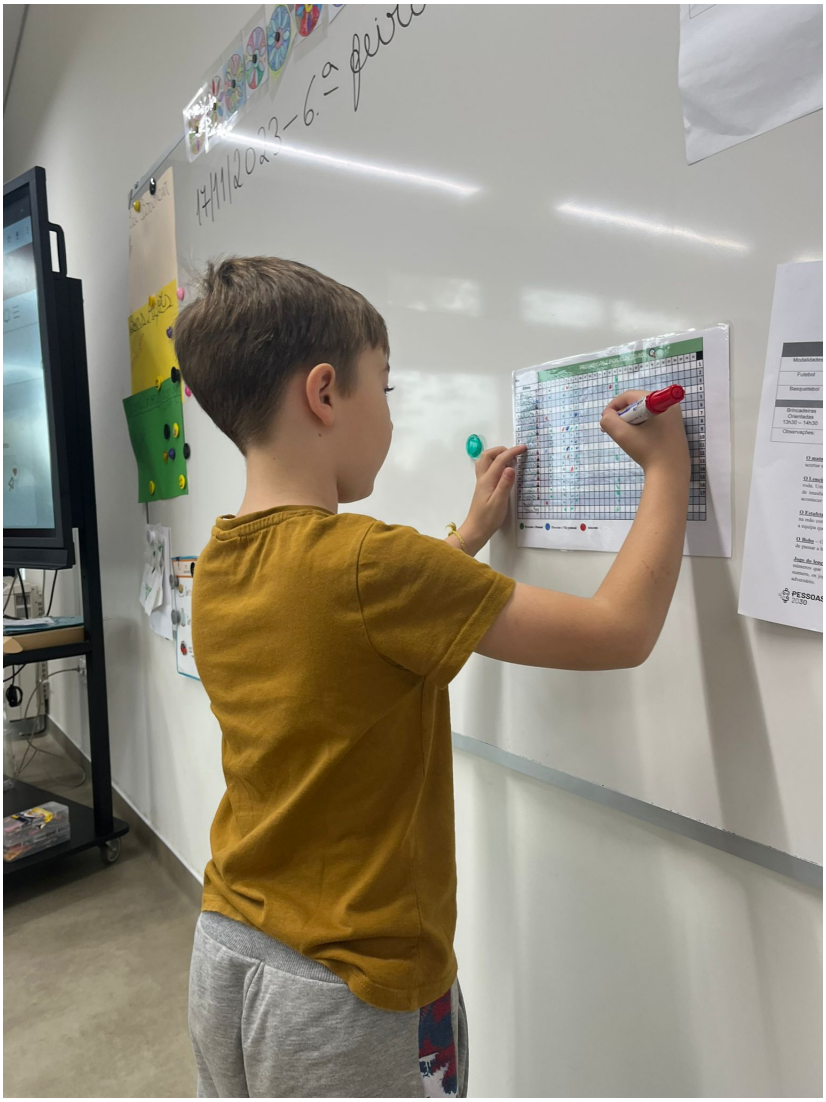


Figure E.1: Student filling in KPI board

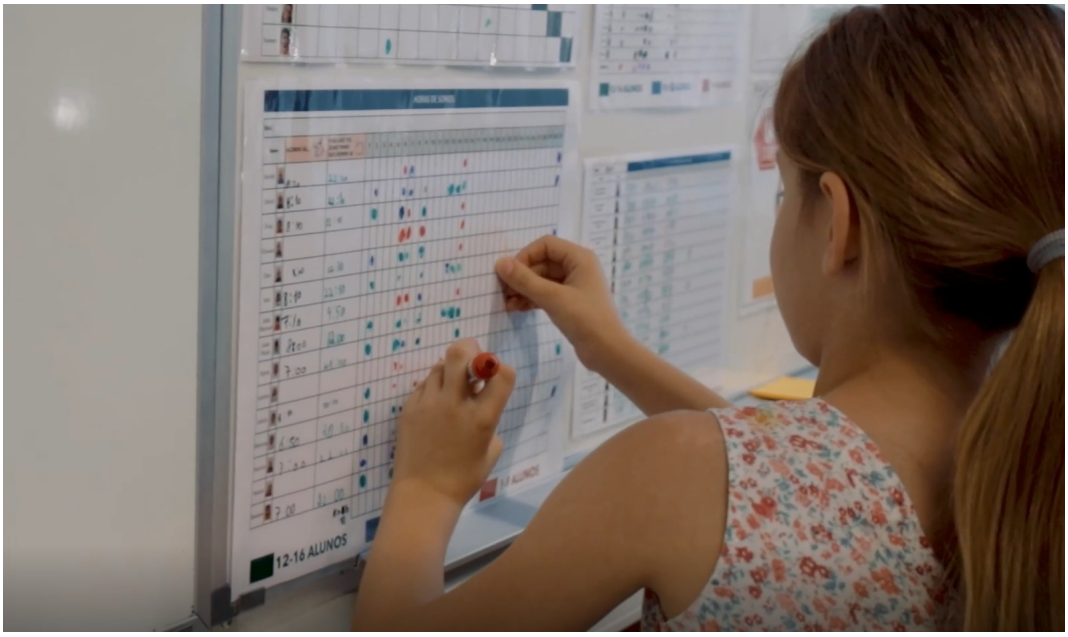


Figure E.2: Student filling in KPI board (2)



Figure E.3: Daily Kaizen Dashboard



Figure E.4: Daily Kaizen Dashboard (2)

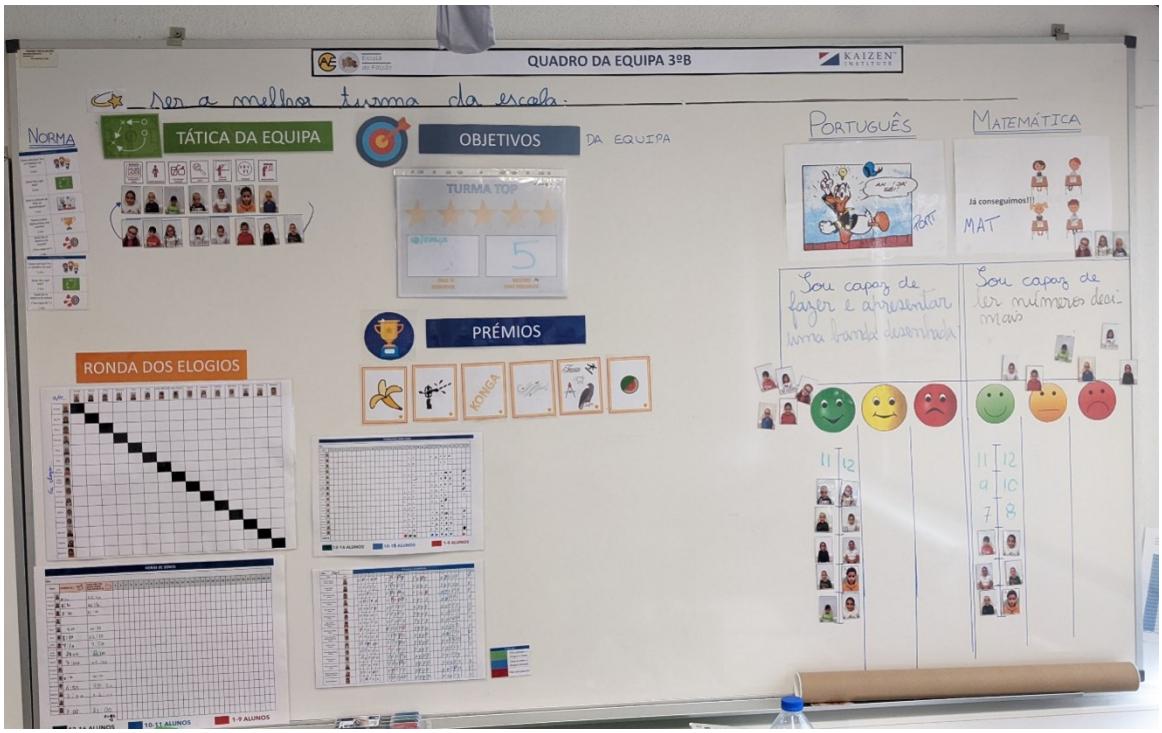


Figure E.5: Daily Kaizen Dashboard (3)