iscte

INSTITUTO UNIVERSITÁRIO DE LISBOA

Student Assessment in Portugal: Impacts of the Covid-19 Pandemic in an Ongoing Process of Change

Gabriel Cipriano

Doutoramento em Políticas de Administração e Gestão Escolar

Orientadora: Doutora Susana da Cruz Martins, Professora Associada, Iscte

Outubro, 2024



Departamento de Ciência Política e Políticas Públicas

Student Assessment in Portugal: Impacts of the Covid-19 Pandemic in an Ongoing Process of Change

Gabriel Cipriano

Doutoramento em Políticas de Administração e Gestão Escolar

Orientadora: Doutora Susana da Cruz Martins, Professora Associada, Iscte

Outubro, 2024



Departamento de Ciência Política e Políticas Públicas

Student Assessment in Portugal: Impacts of the Covid-19 Pandemic in an Ongoing Process of Change

Gabriel Cipriano

Doutoramento em Políticas de Administração e Gestão Escolar

Júri:

Doutor Luís Capucha (Presidente), Professor Catedrático, Iscte Doutora Maria Assunção Flores Fernandes, Professora Associada com Agregação, Universidade do Minho Doutor Manuel Gomes, Vogal do Conselho Diretivo, Instituto de Avaliação Educativa (IAVE) Doutor Domingos Fernandes, Professor Catedrático, Iscte Doutora Susana da Cruz Martins, Professora Associada, Iscte

Outubro, 2024

Declaration

In accordance with the General Regulatory Norms for Doctoral Programs at Iscte (Regulation No. 353/2016), this thesis is an original work specifically prepared and submitted to partially fulfil the requirements for the Doctor of Philosophy degree in Policies of School Administration and Management. Therefore, this thesis is fully authored by the candidate, and any use of contributions or external texts is duly referenced.

fundação para a Ciência e a Tecnologia

This thesis was made possible with the financial support of the *Fundação para a Ciência e Tecnologia*, through the award of a doctoral research studentship with reference number 2020.05847.BD. The author expresses his gratitude for this support, without which the research would not have been feasible. More information available at <u>https://doi.org/10.54499/2020.05847.BD</u>.

It was the year of 1987 when my grandmother and I went to school for the first time. At the time, I remember rehearsing together, in very crooked handwriting, the writing of our name, as if trying to dance a delicate ballet choreography for a pen on a sheet of paper. The sixty years of age that separated us did not diminish the joy we both felt of, for the first time, being able to obtain our identity card with our name written by ourselves. We then began to admire with pride and vanity that piece of yellow plasticized paper, with a photograph and a smudge of our fingerprint which, for us, was an official diploma, the very high recognition from the Portuguese Republic that we knew how to read and write. As time went on, the writing of my name became less crooked and rounder. My grandmother, on the other hand, between the chores on the farm and the sales at the market, began to have a less and less perceptible handwriting, which she ended up forgetting. Later, she wanted to renew her identity card stating that she was illiterate again, but she was denied that option:

- How does one forget how to write one's name?

Thus, having effectively forgotten the dance of her own name for a pen on a sheet of paper, whenever she needed to sign something, I would perform my choreography, in her name, with her name, in crooked handwriting to be credible. As I present this thesis to conclude the highest degree of my educational pathway, I cannot help but remember my first companion in my schooling process. We both had the same opportunity to learn to read and write. We both had the dream of writing to the whole world. We both had the aptitude to make it this far. However, my grandmother's time was not my time. Since this was not her time, as in the old times, I write her name once again, so that her name is never forgotten and forever remembered.

> To the memory of my grandmother, my first companion in the dance of the letters, Bárbara da Conceição.

Acknowledgments

In the year in which we celebrate the 50th anniversary of Portuguese democracy, I must first acknowledge all those man and woman who dreamed the "25 de Abril de 1974" and created the opportunities that allowed me to reach this point. Unlike my grandmother, I was fortunate enough to attend school until the age of 18 completely free of charge. In higher education, I began my academic pathway with the support of a social scholarship, and later I had the opportunity to develop my doctoral work, presented here, thanks to a doctoral research studentship provided by the *Fundação para a Ciência e Tecnologia*. While much remains to be done to fulfil "April's dream", I express my deepest gratitude to the Portuguese Republic and the European Union for making me believe that I was born and raised in the right place, at the right time.

Second, I would like to express my gratitude to all my professors at Iscte. I have learned a great deal from each of them, and they are the reason why Iscte is truly "*a space to grow*". In particular, I want to extend a warm acknowledgment to my supervisor, Professor Susana da Cruz Martins, who has been tireless in her support since the day I proposed her as my master's supervisor. For this reason, I feel that the doctoral thesis presented here is not solely mine; it belongs to us. Additionally, I cannot forget to thank Professor Helena Carvalho for her invaluable lessons in statistics, for the opportunities at Iscte, and for always encouraging me to raise my bar: *"You only fall from the top"*.

Third, I would like to extend my gratitude to all those who contributed to the writing of this thesis. This includes the *Conselho Nacional de Educação* (CNE), particularly Professor Manuel Miguéns, for sharing a part of the data analysed in this thesis; the 32 school' head teachers who generously participated in interviews; the *Instituto de Avaliação Educativa* (IAVE), especially Professor Luís Santos, for his willingness to meet with me and grant an interview; and the more than 2 600 teachers who spared a few minutes of their valuable time to participate in a survey questionnaire.

Fourth, I want to extend a heartfelt thank you to my colleagues Bruno Ferreira, Cândido Fernandes, Diogo Paiva, and Rolando Volzone for the friendship that accompanied our academic journeys at Iscte, and to Dan-Anders Normann for his companionship within the Postgraduate Student and Early Career Researcher Network of the Association for Educational Assessment – Europe.

Last but certainly not least, I want to acknowledge my life partner, António. None of this would have been possible without you. None of this would make sense without you.

Resumo

Em março de 2020, considerando os alertas da Organização Mundial de Saúde e perante o real agravamento dos riscos da pandemia Covid-19 para a saúde pública, o governo português determinou o encerramento do espaço físico das escolas de todo o país. Estabelecido o ensino remoto de emergência, os desafios que até então existiam na avaliação das aprendizagens tornaram-se desafios ainda mais complexos. Por isso mesmo, todo este período de confinamento provocou, não só em Portugal como um pouco por todo o mundo, uma forçada reflexão sobre os propósitos, práticas e uso da avaliação das aprendizagens. Para investigar e analisar alguns dos impactos desta pandemia no domínio da avaliação educacional em Portugal, procurou-se, ao longo dos últimos quatro anos, compreender o enquadramento legal que sustentou a realização da avaliação das aprendizagens durante a pandemia, bem como as suas implicações nos resultados e percursos escolares dos alunos, na adaptação e no volume de trabalho percecionado pelos professores, na relação estabelecida entre a avaliação interna e externa das aprendizagens, e nas políticas e práticas de avaliação atuais. Para dar resposta a estes objetivos, numa primeira fase, recorreu-se à análise documental e à análise secundária de dados. Posteriormente, realizou-se um largo conjunto de entrevistas para melhor compreender o fenómeno em estudo e, por último, foi disseminado um inquérito por questionário, com o objetivo de fazer considerações e ilações quantitativas. Os resultados do estudo que aqui se apresenta apontam para que tenha havido uma efetiva mudança nas políticas e práticas de avaliação das aprendizagens durante a pandemia, com implicações nas políticas e práticas de avaliação atuais; contribuindo para uma importante reflexão sobre os problemas que afetam a educação em Portugal.

Palavras-chave

Políticas de administração e gestão escolar; Políticas de avaliação das aprendizagens; Práticas de avaliação das aprendizagens; Pandemia Covid-19; Ensino remoto de emergência; Avaliação de emergência.

Abstract

In March 2020, considering the warnings of the World Health Organisation and facing a real aggravation of the Covid-19 pandemic risks to the public health, the Portuguese government ordered the closure of physical school spaces throughout the country. With the establishment of emergency remote teaching, the previous existing challenges in assessment of learning became even more complex challenges. Therefore, this entire confinement period has sparked, not only in Portugal but also around the world, a forced reflection on the purposes, practices, and uses of assessments. During the last four years, to study and to understand some of the impacts of this pandemic on the domain of educational assessment in Portugal, it was sought to comprehend the legal framework that supported assessment of learning during this period, as well as its implications on students' results and schooling pathways, teachers' adaptation and perceived workloads, the relationship between internal and external assessment of learning, and current assessment policies and practices. To fulfil these objectives, in a first phase, a documentary analysis and secondary data analysis were performed. Subsequently, a wide range of interviews were conducted to better understand the phenomenon under study, and lastly, a survey questionnaire was disseminated, aiming to draw quantitative considerations and conclusions. The results of the study presented here point to an effective change in the assessment of learning policies and practices during the pandemic, with implications for the current assessment of learning policies and practices; contributing to an important reflection on the problems affecting education in Portugal.

Keywords

Policies of school administration and management; Assessment of learning policies; Assessment of learning practices; Covid-19 pandemic; Emergency remote teaching; Emergency assessment.

Index

Acknowledgments	vii
Resumo	ix
Abstract	xi
Index	xiii
List of Figures	xix
List of Tables	xxi
List of Acronyms	xiii

Intr	oduction	1
	Initial Considerations and Analytical Model Proposal	1
	Objectives and Research Design	3
	Thesis Structure	4
	Concluding Remarks	7

hapter 1. Evaluation and Assessment in the Portuguese Compulsory Education System	9
1.1. Evaluation	9
1.1.1. The Nature of Evaluation 1	0
1.1.2. Evaluation Perspectives 1	1
1.1.3. Evaluation Purposes 1	4
1.1.4. Evaluation Quality 1	6
1.2. Educational Assessment 1	7
1.2.1. Student Assessment Purposes 1	8
1.2.2. Assessment of Learning, for Learning and as Learning	0
1.2.3. Audiences 2	2
1.3. The Portuguese Education System 2	3
1.3.1. Governance 2	3
1.3.2. Compulsory Education System Structure 2	4
1.3.3. Evaluation and Assessment Framework in Portugal	6
1.4. Policies for Student Assessment in Compulsory Education 2	8

	1.4	4.1. The PASEO	29
	1.4	4.2. The Decree-Law No. 55/2018	32
	1.4	4.3. The Decree-Law No. 54/2018	37
	1.4	4.4. The MAIA Project	38
	1.4	4.5. The DAVE Project	40
1.5		Higher Education Access Regime	42
1.6	5.	Chapter Final Considerations	43

Chapter 2. Student Assessment During the Covid-19 Pandemic	. 45
2.1. The Covid-19 Pandemic Context	. 45
2.1.1. Timeline of the Context	. 46
2.1.2. Emergency Remote Teaching	. 48
2.2. The Decree-Law No. 14-G/2020: An Emergency Assessment Policy	. 50
2.2.1. Student Internal Assessment Policy in ERT	. 51
2.2.2. Student External Assessment Policy in ERT	. 52
2.3. Early Impacts of ERT on Student Assessment	. 53
2.3.1. ERT Impacts on Internal Assessment	. 53
2.3.2. ERT Impacts on External Assessment	. 55
2.3.3. ERT Impacts on Grade Repetition	. 59
2.3.4. ERT Impacts on Digital Transition	. 61
2.4. Chapter Final Considerations	63

hapter 3. Research Design	
3.1. Taxonomy Development and Approach Assignment	65
3.1.1. Internal Assessment Practices in ERT	66
3.1.2. Cancellation of External Assessments	67
3.1.3. Grade Repetition Beliefs	69
3.1.4. DAVE's Implementation	70
3.1.5. Research Framework	71
3.2. Sub-study 1: Internal Assessment Practices in ERT	73
3.2.1. Method for Data Analysis	74
3.2.2. Database	76
3.2.3. Procedures for Data Collection	77
3.2.4. Participants	78

3.3.	Sub-study 2: Cancellation of External Assessments	. 80
3	3.3.1. Interviewing	. 81
	3.3.2. Instrument A Development – Interview Protocol	82
3	3.3.3. Sampling and Data Analysis	83
3	3.3.4. Participants	85
3.4.	Sub-study 3: Grade Repetition Beliefs	89
	3.4.1. Sample A	89
3	3.4.2. Sample B	89
3	3.4.3. Instrument B Development – Survey Questionnaire	90
3	3.4.4. Sampling Processes for Data Collection	91
	3.4.5. Participants	93
3	3.4.6. Methods for Data Analysis	95
3.5.	Sub-study 4: DAVE's Implementation	96
3	3.5.1. Instrument C Development – Interview Protocol	96
	3.5.2. Participants and Data Analysis	97
3.6.	Ethical Issues	98
3.7.	Chapter Final Considerations	99

Chapter 4. Assessment of Learning During the 1 st Wave of Covid-19 Pandemic with	a Glance at Special
Education	101
4.1. Hypotheses Formulation	101
4.1.1. Teachers' Adaptation to ERT and Perceived Workloads	101
4.1.2. Instruction and Assessment During ERT (The Mediator)	102
4.1.3. Changing Assessment Methodologies During ERT (The Moderator)	105
4.1.4. Assessment in Special Education During ERT	106
4.2. Method	107
4.2.1. Measures	107
4.2.2. Data Analysis Procedures	109
4.3. Results	109
4.3.1. Preliminary Analysis	109
4.3.2. Hypotheses Testing Considering all Curricular Areas	110
4.3.3. Complementary Analysis: Special Education Area	112
4.4. Discussion	114
4.5. Chapter Final Considerations	117

Chapter 5. The Washback Effect of External Assessments and the Covid-19 Pandemic: Perspectives	of
Head Teachers in Portugal 1	.19
5.1. External Assessments 1	.19
5.1.1. Consequences of External Assessment Tests in Education Systems 1	.20
5.1.2. Consequences of External Assessment Tests in Portugal 1	.21
5.1.3. Research Opportunities and Questions 1	.22
5.2. Method 1	.23
5.3. Results 1	.23
5.3.1. Preliminary Analysis 1	.23
5.3.2. Inequalities in ERT Provision1	.23
5.3.3. Cancelation of External Assessment During the Covid-19 Pandemic 1	.24
5.3.4. External Assessment Reintroduction 1	.25
5.4. Discussion 1	.30
5.5. Chapter Final Considerations 1	.33

Chapter 6	. Examining the Effects of the Covid-19 Pandemic and Professional Development on Teach	ners'
Grade Re	petition Beliefs and Practice	135
6.1.	Grade Repetition	135
6.	1.1. Assessment and Grading During the Covid-19 Pandemic in Portugal	136
6.	1.2. Assessment Literacy and Professional Development During the Pandemic	137
6.	1.3. Concepts Clarification for the Study	137
6.2.	Method	138
6.	2.1. Measures	138
6.	2.2. Procedures and Data Analysis	139
6.3.	Results	139
6.	3.1. Pre and Post Covid-19 Teachers' Grade Repetition Beliefs	139
6.	3.2. Understanding the Effect of Participation on the MAIA Project	142
6.	3.3. Understanding the Effects of Sex, Age, Qualification, School Sector and Teaching L	evel
01	n Teachers' Grade Repetition Beliefs	143
6.	3.4. Responsibility for Educational Success and Grade Repetition	145
6.	3.5. The MAIA Project Effects on Classroom Assessment Practices	146
6.4.	Discussion	147
6.5.	Chapter Final Considerations	149

Chapter 7.	Introducing Technologies into National Large-Scale Testing: Are we Ready?	151
7.1.	Large-Scale Testing	151
7.1	.1. Introducing Technologies into National Large-Scale Testing	152
7.1	.2. Window of Opportunity for Computer-Based Large-Scale Testing in Portugal	154
7.1	.3. Research Opportunities and Questions	155
7.2.	Methods	156
7.3.	Results	156
7.3	.1. Head Teachers' Concerns Regarding the Implementation of DAVE	156
7.3	.2. Concerns Covered in DAVE's Design and Implementation	160
7.3	.3. Teachers' ICT Use in the Post Pandemic Period	161
7.3	.4. Teachers' Acceptability of External Assessments	164
7.3	.5. Teachers' Acceptability of DAVE	165
7.4.	Discussion	166
7.5.	Chapter Final Considerations	168

Conclusion		171
	Summary of Achievements	171
	Contributions to Student Assessment Policy	173
	Limits and Challenges for Research	176
	Final Considerations	178

References	179
Other Sources	191
Legislation	193

pendixes 19	95
Appendix A. Head Teachers' Interview Script	I
Appendix B. Teachers' Questionnaire	۷
Appendix C. IAVE's president interview script X	(11)
Appendix D. Ethical Issues XV	٧II
Appendix E. Mann-Whitney U tests from Chapter 6 X	XI

List of Figures

Figure 1. Analytical model	2
Figure 2. Structure of the National Education System	25
Figure 3. Conceptual Framework for the Students' Profile by the End of Compulsory Schooling	32
Figure 4. External Assessment in the Portuguese Compulsory School System	34
Figure 5. Number of students enrolled in secondary education and number of exams taken per	
academic year (2018-2020)	55
Figure 6. Secondary education conclusion rate from 2010 to 2022	56
Figure 7. Grade repetition and dropout rates, by academic year and grade level (%)	60
Figure 8. The process of deduction for sub-study 1	73
Figure 9. Simple mediation model with a single mediator variable M	74
Figure 10. Simple moderation model with a single moderator variable W	74
Figure 11. Variants of a conditional process conceptual model	75
Figure 12. Teachers' distribution by NUT III	78
Figure 13. Teachers with coordination functions' age (years)	79
Figure 14. Teachers (%) by curricular area	80
Figure 15. The process of theoretical sampling	83
Figure 16. Number of analytical categories by analysed interview number	85
Figure 17. Head Teachers and schools' distribution (n) by NUT II	88
Figure 18. Sample B Teachers' distribution by NUT III	93
Figure 19. Conceptual model for Hypothesis 1	102
Figure 20. Conceptual model for Hypothesis 2	104
Figure 21. Conceptual model for Hypothesis 3	105
Figure 22. Conceptual model for Hypothesis 4	106
Figure 23. Moderation effect of changed assessment methodologies in the relation between	
teachers' adaptation and instruction and assessment	112
Figure 24. Head teachers' profiles regarding external assessment reintroduction	126
Figure 25. Teachers' grade repetition beliefs and practice, 2019 (sample A) and 2023 (sample B)	
(%)	140
Figure 26. Teachers' post-pandemic assessment practices by sub-sample (%)	146
Figure 27. Head Teachers' Concerns and Profiles regarding DAVE	157

Figure 28. Teachers use of ICT in their classroom practices (%)	162
Figure 29. Degree of Agreement (%) with Digital Changes Due to the Covid-19 Pandemic	163
Figure 30. Paper-Based and Digital Testing in Classroom Assessment (%)	163
Figure 31. Agreement with PA existence (%) within the subgroup of teachers who agree with, at	
least, one PA	164
Figure 32. Agreement with exams existence (%) within the subgroup of teachers who agree with,	
at least, one exam	165

List of Tables

Table 1. Main analytical axis of each thesis' chapter	7
Table 2. Formative Evaluation and Summative Evaluation	15
Table 3. Characteristics of Formative and Summative Assessments	19
Table 4. Synthesis of Portuguese Legislation	28
Table 5. Grade repetition and dropout rates, by academic year and grade level (%)	36
Table 6. DAVE implementation plan	41
Table 7. Main events of the Covid-19 pandemic outbreak in Portugal	46
Table 8. Average scores in EN (0-200 points) and year-on-year variation rate (%) from 2015-2020.	57
Table 9. Changes to 2019/20 national examinations due to the Pandemic within OECD countries	58
Table 10. Year-on-year variation of grade repetition and dropout rates, by academic year and	
grade level (pp)	60
Table 11. Research Design Summary	72
Table 12. Teachers with coordination functions by sex and age	79
Table 13. School Characteristics by Cluster Type, Administrative and governance characteristics,	
and Region (n)	. 86
Table 14. School Characteristics by Cluster Type, Administrative and Governance Characteristics,	
and Region (n, N)	. 87
Table 15. Teachers from the compulsory school system by sex, age, qualification, school sector	
and teaching level	. 94
Table 16. Descriptive statistics, reliabilities, and study variables correlations	109
Table 17. Results for mediation considering all curricular areas	110
Table 18. Results for moderated mediation considering all curricular areas	111
Table 19. Results for mediation for special education curricular area	113
Table 20. Results for moderated mediation for special education area	114
Table 21. Distributions comparisons between samples and subsamples	141
Table 22. Distributions comparisons within subgroups	144
Table 23. Teachers' Perspectives on Responsibility for Educational Success and Grade Repetition	
(%)	145
Table 24. Teachers post-pandemic assessment practices' distributions comparisons between	
subsamples	147

List of Acronyms

- AERA American Educational Research Association
- ANPRI [National Association of Informatic Teachers]
- CAM Changed Assessment Methodologies
- CIPP Context, Inputs, Process, and Products
- CNE [National Council of Education]
- DAVE [External Assessment Dematerialization Project]
- DGAE [Directorate General for School Administration]
- DGE [Directorate General for Education]
- DGEEC [Directorate General for Education and Science Statistics]
- DGES [Directorate-General for Higher Education]
- DGS [Directorate-General of Health]
- DL Decree-Law
- DRAEM [Regional Directorate of School Administration of Madeira]
- DREAEA [Regional Directorate of Education and Educational Administration of the Azores]
- EA External Assessment
- EN Exames Nacionais [11th and 12th grades National Exams]
- ERT Emergency Remote Teaching
- EU European Union
- HR Human Resources
- IA Instruction and Assessment
- IAVE [National Institute for Educational Assessment]
- ICT Information and Communications Technology
- IGEC [General Inspectorate of Education and Science]
- ISCED International Standard Classification of Education
- KMO Kaiser Meyer Olkin measure
- LBSE [Education Act]
- MAIA [National project of training, supervision, and research in classroom assessment]
- MAXQDA MAX Qualitative Data Analysis
- ModA Provas de Monitorização da Aprendizagem [Assessments for Learning Monitoring]
- NAE National Academy of Education

- NRC National Research Council
- NUTS Nomenclature of Territorial Units for Statistics
- OCEPE [Curricular Guidelines for Early Childhood Education]
- OECD Organisation for Economic Cooperation and Development
- PA Provas de Aferição [Assessment tests]
- PASEO [Students profile by the end of compulsory education]
- PFC Provas Finais de Ciclo [9th grade National Exams]
- PIRLS Progress in International Reading Literacy Study
- PISA Programme for International Student Assessment
- PRR Portuguese Recovery and Resilience Plan
- RQ Research Question
- SEN Special Educational Needs
- SEND Special Educational Needs and Disabilities
- SGEC [General Secretariat for Education and Science]
- SPSS Statistical Package for Social Sciences
- TA Teachers' Adaptation
- TALIS Teaching and Learning International Survey
- TEIP [Educational Territories of Priority Intervention]
- TIMSS Trends in International Mathematics and Science Study
- UNESCO United Nations Educational, Scientific and Cultural Organization
- UNICEF United Nations Children's Fund
- WHO World Health Organization
- $\mathsf{WL}-\mathsf{Workload}$

Introduction

Initial Considerations and Analytical Model Proposal

Educational assessment is a process for gathering information that can be used to make decisions about students, teachers, curriculum, programmes, and schools, as well as other aspects of education systems and policy (NAE, 2021). This process occurs at different levels of education systems, ranging from the individual student assessment conducted by classroom teachers to national assessments and cross-national comparisons of student achievement carried out by national or international agencies (Kellaghan et al., 2003; Santiago et al., 2012). Although it took many years for educational assessment of education assessment, such assessment remains a central component of educational assessment to inform education policy (Kellaghan et al., 2003).

In March 2020, as a response to school closures caused by the Covid-19 pandemic, Emergency Remote Teaching (ERT) was implemented in most countries worldwide (Hodges et al., 2020). As the Covid-19 virus spread globally, school systems were in crisis management, with education leaders and teachers struggling to deliver continuous instruction through various modes such as in-person, virtual, and hybrid modes (NAE, 2021). In Portugal, the closure of schools and the implementation of ERT led to significant disruptions in teaching methods and learning processes (see, for example, CNE, 2021a; Costa & Baptista, 2023; Flores & Gago, 2020; Flores et al., 2023), with profound implications for students' assessment policy and practice (see, for example, CNE, 2021a; Cooper et al., 2022; NAE, 2021; Nisbet & Shaw, 2022; Panadero et al., 2022; Rodrigues et al., 2022; Seabra et al., 2021).

In this very disruptive scenario, on the one hand, there were legitimate concerns about the use of student assessments without guarantees of validity, reliability, and fairness. On the other hand, teachers needed evaluative information more than ever before to guide classroom learning, and states, schools, parents, and communities needed evidence of how the Covid-19 pandemic was affecting students and their learning (NAE, 2021) to inform and (re)define education policies.

This doctoral thesis, within the scope of the PhD programme on Policies of School Administration and Management at Iscte, aims to provide a comprehensive examination of the Covid-19 pandemic impacts on student assessment policy and practice in Portugal, with a particular focus on the implications of Decree-Law No. 14-G/2020 during the 2019/2020 academic year. Additionally, the study seeks to analyse the effects of implementing this legislation in mitigating the negative consequences of the pandemic. Furthermore, recognizing the central roles of teachers and head teachers in student assessment (Kellaghan et al., 2003), substantial attention is given to teachers' internal assessment practices, to the perceptions and approaches of head teachers towards assessment of learning, and to the interplay between internal and external assessments, as illustrated in Figure 1.



Figure 1. Analytical model

It is important to note that the implementation of public policies typically involves multiple actors and decision-making levels (Souza, 2006), resulting in varied and different interpretations and implementations across different contexts (Ball, 2008). Likewise, educational assessment policies in Portugal also involve various actors and decision-making levels (Santiago et al., 2012), interacting with teachers' and head teachers' own contexts, beliefs, and practices (Cipriano & Martins, 2021), as well as with other national and/or school policies and projects.

During the Covid-19 pandemic, alongside the existing legal framework supporting student assessment, namely the Decree-Law No. 55/2018, the introduction of Decree-Law No. 14-G/2020 coincided with the beginning of other national projects impacting student assessment practices. Therefore, in addition to examining the implications of Decree-Law No. 14-G/2020 on assessment practices, this thesis also aims to investigate the interaction effect with two important national assessment projects: the Training, Supervision and Research in Classroom Assessment (MAIA) project, aimed at enhancing teachers' assessment literacy and practices through professional development (Fernandes, 2021; MAIA project, 2023); and the Dematerialization of External Assessment (DAVE) project, which seeks to integrate and dematerialize all procedures related to external assessment processes (IAVE, 2022), as also depicted in Figure 1.

For this reason, the thesis title "Student Assessment in Portugal: Impacts of the Covid-19 pandemic in an ongoing process of change" not only aim to reflect the dynamic crisis management and unpredictability experienced by different education actors during the pandemic, but also acknowledges the ongoing evolution resulting from the interaction of various assessment projects with the pandemic management and the implementation of Decree-Law No. 14-G/2020 in diverse contexts.

* * *

Objectives and Research Design

The objectives of this thesis are as follows:

- To investigate whether the enactment of Decree-Law No. 14-G/2020 eased the challenges faced by teachers in their teaching and assessment practices, with a particular focus on teachers of vulnerable groups such as students with Special Education Needs and Disabilities (SEND).
- 2. To explore the implications of Decree-Law No. 14-G/2020 on the perception of school head teachers regarding the relationship between internal and external assessment of learning and its impact on school management.
- 3. To identify the impacts of the Covid-19 pandemic on teachers' beliefs towards grade repetition, examining the interaction effect with the MAIA project.
- 4. To study the extent to which the Covid-19 pandemic context has facilitated the digital transition in education, and the challenges faced by school communities in the implementation of the DAVE project.

In terms of research planning to meet these objectives, to be able to answer different research questions of different nature with different objectives (Bryman, 2012:640), it was decided to address each of the four research objectives independently to allow for in-depth exploration of each objective. As a result, the methodology designed and adopted for the research project, referring to the process where different methods, techniques, and materials are applied and used (Oliveira, 2005), was done through the use of mixed methods, combining both quantitative and qualitative data, as well as different stakeholders' perspectives. To this end, at first, a secondary data analysis was conducted. Secondly, a comprehensive set of interviews was carried out to gain deeper insights into the phenomenon under study. Finally, a survey questionnaire was disseminated to gather data to draw

quantitative conclusions. Consequently, the strategy adopted led to the production of four distinct sub-studies, with each sub-study focusing on a specific objective.

After being defined, the research project that supports this thesis received funding approval from the *Fundação para a Ciência e Tecnologia* (reference number 2020.05847.BD). It was then endorsed in a public examination session by the scientific board of the Policies of School Administration and Management PhD programme at Iscte and obtained ethical approval from the Ethics Council of Iscte – IUL (reference number 121/2021).

Throughout the research, the preliminary results of each of these four sub-studies have been presented at national and/or international scientific events, and they are either published or in the process of being published in international scientific journals or in the proceedings of scientific events.

* * *

Thesis structure

To report the research in this thesis, Chapter 1 discusses the main theoretical issues of educational evaluation and assessment. Additionally, it describes and outlines the Portuguese Compulsory Education System, as well as the Evaluation and Assessment legal framework with a particular focus on national policies for student assessment.

In Chapter 2, a brief description of the Covid-19 pandemic context is provided, and the impacts of ERT on assessment practices in Portugal during the first wave of the Covid-19 pandemic are examined through the analysis of several studies and statistical indicators. Additionally, the implications of Decree-Law No. 14-G/2020 during this initial phase of the pandemic are analysed in terms of the continuity or discontinuity of existing assessment practices. The chapter also highlights changes in the purposes and uses of assessment during this period, along with their main implications for grade repetition and students' educational pathways. The research featured in Chapter 2 was presented at the 22nd annual conference of the Association for Educational Assessment – Europe (online) in November 2021, under the theme "Assessment for Changing Times: Opportunities and Challenges". In the same month of November 2021, the research carried out in this chapter was also presented at the "X Symposium on School Organization and Management" (online), organized by the University of Aveiro, focusing on "The Management of Uncertainty in Education: (Dis)continuities and Challenges". Later, this researched was published in Portuguese as a peer-reviewed article in the proceedings book of the latter scientific event (see Cipriano & Martins, 2023).
In Chapter 3, the thesis research design is defined. Considering the formulated research questions and objectives, the research approach for each sub-study – whether qualitative, quantitative, or mixed methods – is outlined, including clarifications of data sources, sampling processes, inclusion/exclusion criteria, and participant characteristics. Additionally, the instruments developed to collect primary data are explained, along with data analysis techniques and ethical considerations.

In Chapter 4, regarding sub-study 1, a secondary data analysis is conducted with data collected from 3 932 teachers in Portugal as part of a study carried out by the Conselho Nacional de Educação (National Council of Education) in 2019/20 (see CNE, 2021a). Using a moderated mediation statistical model, the analysis conducted in this chapter reveals that the relationship between teachers' adaptation to ERT and their perceived workloads is mediated by instruction and assessment practices. Furthermore, it is demonstrated that this indirect effect is moderated by adjustments made in assessment methodologies. Thus, adapting assessment methodologies emerged as a crucial strategy for facilitating better adjustment to ERT. Moreover, when considering special education teachers only, it was observed that changing assessment methodologies did not moderate the relationship between teachers' adaptation to ERT and instruction and assessment, showing that the difficulties experienced by special education teachers went beyond changing assessment methodologies. The data analysis presented in this chapter was conducted as part of the postgraduate course in data analysis for social sciences, under the guidance of Professor Helena Carvalho. The research conducted for this chapter writing is currently in the process of being published as an article in an indexed scientific journal, coauthored with Professor Helena Carvalho and Professor Susana da Cruz Martins. Additionally, this work was presented at the 35^{ème} colloque de l'Association pour le Développement des Méthodologies d'Évaluation en Éducation in January 2024 (held in Braga, Portugal), under the theme "L'évaluation face aux défis de la diversité et de l'inclusion: entre normes et différenciations".

In Chapter 5, containing sub-study 2, the perspectives of Portuguese head teachers about the impact of the temporary suspension of external assessment of learning during the Covid-19 pandemic and its implications for school management are explored. Additionally, the chapter discusses, from the viewpoint of Portuguese head teachers, the external assessment model that should be implemented in the post-pandemic context. To achieve this, 32 semi-structured interviews were conducted with head teachers from across the country. The results reveal that the pandemic has prompted a reconsideration of the purposes and format of external assessments, highlighting varying perspectives on their reintroduction in the Portuguese context. Furthermore, the chapter examines the negative effects of external high stakes assessments, particularly for low-achieving students and school management. The research carried out in this chapter was presented at the *34^{ème} colloque de l'Association pour le Développement des Méthodologies d'Évaluation en Éducation* in April 2023 (held in Mons, Belgium), under the theme "Évaluation des apprentissages: continuités et ruptures". It was

later presented in a webinar organized by the *Association for Educational Assessment – Europe* (online) as part of the *Postgraduate Student and Early Career Researcher Network Webinars*. Additionally, it is currently undergoing publication in an indexed scientific journal, co-authored with Professor Susana da Cruz Martins.

In Chapter 6, encompassing sub-study 3, the impact of the Covid-19 pandemic on Portuguese teachers' beliefs towards grade repetition is studied, as it was observed a substantial drop on grade repetition rates during the 2019/20 school year. It also examines the interaction effect of professional development in assessment literacy, more specifically, the MAIA project, on these beliefs. To this end, a survey involving 2 673 teachers was conducted following the challenge proposed by Cipriano & Martins (2021). This analysis, based on data from two independent samples from two different time periods, reveals that while the Covid-19 pandemic had minimal influence on teachers' grade repetition beliefs, the MAIA project had a significant effect. These findings underscore the importance of teachers' professional development in enhancing classroom assessment practices. The research conducted on this chapter was presented at the 65th World Assembly – International Council on Education for Teaching in June 2024, (held in Braga, Portugal), under the theme "Enhancing the Teaching Profession Through Quality Teacher Education: Purpose, Policy, and Practice in Times of Teacher Shortage". Additionally, it is currently undergoing publication in an indexed scientific journal, co-authored with Professor Susana da Cruz Martins.

In Chapter 7, concerning sub-study 4, the concerns within school communities regarding the implementation of the DAVE project in the post-pandemic context are examined. To identify and analyse these concerns, an additional question was incorporated into the 32 semi-structured interviews conducted with head teachers on mainland Portugal: "Are schools ready for the implementation of DAVE? Why?". Furthermore, to assess the extent to which these concerns were considered and addressed in the design and implementation of DAVE, a supplementary interview was conducted with the President of IAVE. Additionally, the survey questionnaire involving 2 673 teachers included questions about teachers' agreement with the implementation of DAVE, and teachers' ICT use in the classroom. The results revealed divergent opinions among teachers and head teachers. While some believe that schools are ready for DAVE implementation and support it, a significant proportion of head teachers expressed concerns about the lack of investment prior to implementation and noted widespread teacher resistance. The research conducted for this chapter writing was presented in November 2023 at the 24th annual conference of the Association for Educational Assessment – Europe (held in Valletta, Malta), under the theme "Assessment Reform Journeys: Intentions, Enactment, and Evaluation". Additionally, the research is also published as a peer-reviewed article in the Education Policy Analysis Archives, co-authored with Professor Susana da Cruz Martins (see Cipriano & Martins, 2024).

Lastly, a conclusion is made, where key findings are revisited with the identification of research limitations and new challenges for the future of educational assessment.

To better summarise this thesis structure, Table 1 recaps the main analytical axis of each chapter:

		Assessm	ent level		Т	ime Perio	d		Thema	tic focus	
Chapter	Assessment Policy	Assessment Practice	Internal Assessment	External Assessment	Before Covid-19	During Covid-19	Post Covid-19	Workloads & SEND teachers	Washback effect	Grade repetition & MAIA project	DAVE project
Chapter 1	х		х	х	х						
Chapter 2	х		х	х		х					
Chapter 3					RES	EARCH DE	SIGN				
Chapter 4	х	х	х			х		х			
Chapter 5	х	х		х		х			х		
Chapter 6	х	х	х		х		х			х	
Chapter 7	х			х			х				х

Table 1. Main analytical axis of each thesis' cho	ıpter
---	-------

* * *

Concluding Remarks

To conclude this introduction and thesis framework, it is believed that this research makes significant contributions to scientific knowledge, particularly in the field of educational policy. By providing a deep and integrated analysis of the impacts of Decree-Law No. 14-G/2020, an emergency policy with profound implications for schools, teachers, and students, this thesis prompts a critical reflection on the role of public policies in the real life of school communities, namely, in emergency contexts.

Furthermore, this thesis contributes to the field of educational assessment by examining the purposes, practices, and implications of assessments, particularly in light of the unprecedented disruption caused by the Covid-19 pandemic. As the pandemic raised serious concerns about the

validity, reliability, and fairness of assessments, this thesis emphasises the importance of ensuring assessments uphold these fundamental values in an equitable education system.

Finally, this thesis serves as a contribution to furthering knowledge about the disruptive period experienced globally during the Covid-19 pandemic. It captures the memories and challenges faced by individuals and communities, offering valuable insights into this tumultuous time. By documenting and studying this period, the thesis aims to contribute to a deeper understanding of its impact on Portuguese society.

CHAPTER 1

Evaluation and Assessment in the Portuguese Compulsory Education System

Chapter Summary:

Since evaluation and assessment are complex processes that require a thoughtful reflection on ontological, epistemological and methodological aspects, the first two sections of this chapter explore issues related to the nature, foundations and methods of these processes. The third section of this chapter outlines the Portuguese compulsory education system, while the fourth section presents its evaluation and assessment framework, covering most recent developments regarding student assessment policy, as well as two nationwide assessment projects with impact on student assessment practices. Considering the pressure that access to higher education places on compulsory education, a brief section addressing higher education access regime is also included.

1.1. Evaluation

Practices and reflections on practices are insufficient to ensure an understanding of what we do. Practice is what we do in order to achieve an intended result or outcome. Theory provides the coherence and the logic for the practice, i.e., it provides the basic frame for rationalising empirical research, and it also stands alone as a logical and coherent rationalisation. Therefore, theory links practice and empirical research into a coherent narrative (Taras, 2012).

For a Latin language speaker, such as a Portuguese person, both "Evaluation" and "Assessment" translate to "Avaliação". As such, at the beginning of this thesis, evaluation and assessment theory seem to require a clear definition of its nature, foundations, and methods. However, providing a single and clear definition appears to be neither easy nor universally agreed upon. According to Stufflebeam & Shinkfield (1985), many definitions can be offered to the question "What is evaluation?" due to the diversity of conceptual models that exists. As different approaches to the practice of evaluation have emerged over the years, the definition of evaluation has also evolved with these practices (Stufflebeam & Coryn, 2014). Similarly, Guba & Lincoln (1989) emphasize the difficulty in finding a "correct" way to define evaluation, a definition that would settle the debate on how an evaluation is conducted and what its purposes are. Black & Wiliam (2018), revisiting their seminal article from 1998 after a challenge proposed by Perrenoud, refer that many approaches have also been explored within

educational assessment theories and practices. They also refer that any approach to an assessment theory must address all aspects of assessment in an integrated way, even though such approach risks becoming too broad, and a theory of assessment becomes a rather weak "theory of everything". Fernandes (2013) also highlights that realities are, in general, quite complex, integrating several different elements, which make them difficult to describe, analyse, and interpret. Therefore, it is unwise to consider only one type of *avaliação*, a single way of generating knowledge, or one sole process of collecting information about that reality. Given this difficulty in defining *avaliação*, evaluation and assessment, the following subsections of this chapter explore some of its main features.

1.1.1. The Nature of Evaluation

In common sense, for many years, evaluation was associated with assessing achievement against clearly defined objectives or conducting norm-referenced testing (Stufflebeam & Coryn, 2014). However, the field of evaluation is much broader. In our day-to-day lives, as in many disciplines or areas of research, evaluation is used systematically in a variety of contexts and situations: from evaluating a consumer's level of satisfaction with a product; evaluating an employee's performance within an organization; evaluating the effects of new drugs in the pharmaceutical industry; evaluating an athlete's performance in a sports competition; evaluating the implementation of a public policy; or even evaluating things as simple and mundane as deciding what to wear when waking up or pondering the advantages of going to supermarket A or B (Clarke, 1999). In fact, the act of evaluating is as old as any practice itself, as it constitutes an integral part of any practice (Scriven, 1994). Therefore, evaluation is a social practice, necessarily carried out by individuals, which can be conducted in a wide variety of contexts and scientific fields, in a more intentional or less intentional way, more formally or less formally, but it can never be separated from its human dimension because it is, in its essence, a social practice.

As a social practice that spans many areas, Michael Scriven (1994), in his article *"Evaluation as a Discipline"*, proposes a transdisciplinary approach to evaluation, where it is constituted as its own discipline, serving a wide variety of fields, i.e., with pure subject matter, beyond its applied areas. This conception of evaluation as a transdiscipline gives to evaluation unique characteristics in terms of its scope, as well as its epistemological, methodological, and political dimensions (Scriven, 1994: 160-161), thereby enabling the study, deepening, and development of evaluation concepts and practices.

At the outset, it seems to be important to distinguish between formal and informal evaluation based on how they are conducted. Formal evaluation applies scientific procedures, which collect and analyse information related to the object being evaluated (the evaluand object), clearly specifying the criteria used, the data sources referenced, and presenting the evidence that forms the evaluative judgment (Clarke, 1999; Fernandes, 2013). Whether we are conducting a systematic formal evaluation, usually carried out by a professional evaluator, or an informal subjective evaluation, the kind that we all perform in our daily lives, the act of evaluating involves making a judgment of value or merit about something (Clarke, 1999). This conception of evaluation is based on the derivation of the word 'value' and indicates that evaluation involves a value judgment regarding the merit or worth of an evaluand. According to this perspective, if an evaluation does not indicate how good or bad something is, it is not truly an evaluation. However, Stufflebeam and Shinkfield highlight that some authors refute this definition of evaluation. Authors that refute this definition believe that focusing an evaluation on a judgment inevitably leads to subjective interpretations, giving undue power to the evaluators (Stufflebeam & Shinkfield, 1985: 4). Thus, when an evaluation involves forming a judgment, the evaluation becomes inherently subjective, with evaluators being aware that they will inevitably influence what they are evaluating. On the other hand, when evaluators seek the truth through an evaluation that is as objective as possible, they aim to assume a neutral and detached position regarding the objects being evaluated (Fernandes, 2010, 2019). However, since evaluation is a social practice, it can never be entirely neutral or free from its human dimension. This fact, concerning the evaluator's positioning, which is more or less present but never truly absent, highlights that an evaluation is eminently (and inevitably) subjective and, as a result, cannot be an exact science.

Since evaluation is not completely exact, there is always some error associated, making it impossible to produce fully accurate results. Instead, the results can only be plausible, credible, and useful. However, it is important to emphasize that even though an evaluation is inherently subjective, with associated errors and without exact results, this does not mean it cannot be rigorous. While evaluators cannot be completely neutral, they should strive to be as impartial as possible (Fernandes, 2019b). Later, when discussing the quality of the evaluation, this issue of the evaluator's positioning will be revisited and further developed.

1.1.2. Evaluation Perspectives

When looking at evaluation, there are various perspectives and approaches that can be taken.

The evaluation perspectives are nothing more than prescriptions for evaluating programs or policies that have been suggested by their respective proponents over the years and, as such, they are deeply rooted in their experiences, their conceptual systems, and also in their ideologies and views of the world and society (Fernandes, 2019b:2).

One of the approaches or perspectives on evaluation that can be taken, relates to the positioning of the evaluator (more or less present), which was just explored in the previous section when reflecting on the nature of evaluation.

Another "classic" approach to evaluation is the dichotomy between quantitative and qualitative evaluation. In quantitative approaches, the use of procedures to find valid and impartial answers is valued. In this case, there is a clear concern with the impartiality and neutrality of the evaluators. In contrast, qualitative approaches are more focused on a detailed description of reality, aiming to understand and know the processes of its functioning (Fernandes, 2010). Today, mixed approaches are also considered, which seek to incorporate elements from both of the previously mentioned groups. By combining them, the goal is to mitigate the limitations of each approach, thereby making the evaluation more useful.

We can also examine evaluation practices based on their theoretical frameworks, specifically evaluation theories, program theories, and social science theories. Evaluation theories are essentially prescriptive, providing a set of rules, procedures, and recommendations that determine what constitutes a good evaluation. Program theories, on the other hand, identify and select relevant issues that deserve particular attention, adjusting procedures to that reality. Social science theories aim to study phenomena that occur during the development of a given program and analyse the social conditions that are intended to be improved (Fernandes, 2010; Stufflebeam & Coryn, 2014).

At this regard, Alkin & Christie (2012) also sought to systematize various theoretical perspectives on evaluation in their article "*An Evaluation Theory Tree*" into three major groups/branches. In this tree, on one side, the roots of evaluation are grounded in accountability and control. On the other side, the roots of evaluation are grounded in the systematic knowledge of social realities, i.e., social inquiry. Then, the tree's three main branches develop from there:

- 1) The first branch addresses the methodological issues of evaluation and is the first to emerge, with research carried out by Ralph Tyler.
- 2) The second branch, inspired by the work of Michael Scriven, develops issues related to the value of evaluation.
- The third branch explores issues related to the use of evaluation, inspired by the work of Daniel Stufflebeam (Alkin et al., 2012).

Stufflebeam, known for his CIPP model for evaluation (a comprehensive framework for conducting evaluations of programs, projects, personnel, products, organizations, policies, and evaluation systems, providing direction for assessing <u>Contexts</u>, <u>Inputs</u>, <u>Processes and Products</u> (Stufflebeam & Coryn, 2014)), identified 22 evaluation perspectives and organized them into four major groups:

- 1. Pseudo-evaluations (which promote a positive or negative image of a program, regardless of its actual merit or value).
- 2. Quasi-Evaluation, i.e., evaluations guided by questions and/or methods (which begin with questions and then design methods to answer those questions).
- 3. Evaluations focused on improvement and accountability (designed essentially to evaluate and/or improve the merit and value of a given program).
- Evaluations driven by a social agenda (engaging stakeholders as well as experts in the characterization, investigation, and formation of value judgments about programs) (Stufflebeam, 1999; Stufflebeam & Coryn, 2014).

Another theoretical perspective on evaluation was presented by Guba & Lincoln in 1989, in their book *Fourth Generation Evaluation*. In this book, Guba & Lincoln identify three evaluation perspectives and propose a fourth one: 1) evaluation as measurement, 2) evaluation as description, 3) evaluation as judgment, and 4) evaluation as construction:

- 1. Evaluation as Measurement emerged from the measurement of various student attributes with the massification of schools in the 19th century. The primary purpose of the school was to teach "what was true". Children demonstrated their knowledge of these "facts" through tests, which were essentially memory tests at the time. Today, evaluation continues to serve the purpose of measurement. For example, in national exams, students are evaluated to be measured and ranked on a scale, which is then used for access to higher education.
- 2. Evaluation as Description aims at verification. This perspective derives from the first and also views evaluation as measurement, but with defined objectives to verify whether those objectives have been achieved.
- 3. Evaluation as Judgment arose to improve programs from the National Science Foundation and the Office of Education in the United States. Here, the evaluator is seen as someone who describes reality, and, as the name suggests, the final analysis is expressed as a judgment. In this case, an evaluator is chosen for their expertise and knowledge of what is being evaluated, and their evaluation is inherently subjective.

4. Responsive Constructivist Evaluation, proposed by Guba & Lincoln, seeks to address the shortcomings and limitations identified in the previous three perspectives. The term *Responsive* aims to create a new vision for the focus of evaluation, where parameters and boundaries are not defined a priori as in the previous three perspectives. *Constructivist* is used to describe the methodology employed during the evaluation. Thus, responsive constructivist evaluation presents itself as a constructivist perspective where evaluation is conducted for the purpose of improvement (Guba & Lincoln, 1989).

Whether in the organization of perspectives by Alkin & Christie (2012), in the four categories created by Stufflebeam (1999), or in the perspectives identified by Guba & Lincoln (1989), it seems evident that the most current approaches are centred on concerns about the use and purposes of evaluation. Indeed, of the 22 perspectives identified by Stufflebeam, he considers the most promising for the 21st century to be those aimed at political and social intervention to improve programs and public policies. Therefore, it seems to be important to reflect on the purposes of evaluation.

1.1.3. Evaluation Purposes

The most important purpose of evaluation is not to prove but to improve. (Stufflebeam & Shinkfield, 1985: 151)

For Stufflebeam & Shinkfield (1985), the most important aspect of evaluation is not to prove, but to improve. However, the purposes of an evaluation may not be limited to improvement alone. Stufflebeam & Coryn (2014) mention that four main purposes can be considered: 1) Improvement (the most important), 2) Accountability, 3) Dissemination and 4) Enlightenment. Regarding the first two points of these purposes – improvement and accountability – it is important to bring into the discussion the concepts of formative and summative evaluation, created by Michael Scriven in the 1960s. These two concepts are extensively used in educational settings and are further explored in Table 2.

 Improvement – Evaluation aimed at improvement seeks to provide relevant information and high-quality feedback for the development of a service or learning. Providing feedback for improvement is, in essence, the concept of formative evaluation.

- Accountability Evaluation aimed at accountability typically occurs after the development of a product, at the conclusion of a program, or at the end of a cycle. It provides an overall judgment about the value or merit of the evaluated reality, and this is, in essence, the basis of summative evaluation.
- Dissemination Evaluation aimed to help researchers disseminate proven products or practices, thereby aiding consumers in making informed choices. This is a consumer-oriented evaluation.
- Enlightenment evaluators seek to clarify or provide new insights into a given reality through their evaluation. This is an evaluation aimed at clarifying stakeholders, that is, those who may be interested in the evaluation results (Stufflebeam & Coryn, 2014:21).

Descriptor	Formative Evaluation	Summative Evaluation
Purpose	Quality assurance; improvement	Providing an overall judgement of the evaluand
Use	Guiding decision making	Ascertaining accountability for successes and failures; promoting understanding of assessed phenomena
Functions	Provides feedback for improvement	Informs consumers about an evaluand's value (for example, its quality, cost, utility)
Orientation	Prospective and proactive	Retrospective and retroactive
When conducted	During development or ongoing operations	After completion of development
Nature of evaluation plans	Flexible, emergent, responsive, interactive	Relatively fixed, not emergent or evolving
Nature of reports	Periodic, often relatively informal	Containing a cumulative record and assessment of what was done and accomplished
Relationship between formative and summative evaluation	Often forms the basis for summative evaluations and supplements summative evaluations.	Involves compiling, assessing, and building on previously collected formative evaluative information

Table 2. Formative Evaluation and Summative Evaluation

Source: Adapted from Stufflebeam & Coryn, 2014:23

The purposes of an evaluation must be considered when choosing data collection procedures, the type of interaction the evaluator will establish with participants, the stakeholders (those interested in the evaluation results), and the method of disseminating the evaluation results (Fernandes, 2010, 2013). Considering all these variables, the evaluation must also be guided by principles that confer rigor, utility, significance, and social relevance. In other words, that ensure its quality.

1.1.4. Evaluation quality

If the essence of any evaluation process lies in discerning the quality of what is being evaluated (the evaluand object), this implies that the evaluation itself must be of high quality. According to Stufflebeam & Coryn (2014), an evaluation is only of quality if it is guided by:

- a) Integrity and Ethical Adequacy It is of public interest that all evaluations are conducted with integrity and ethical adequacy. An evaluation lacking integrity is not valuable and thus has no utility. On the contrary, an evaluation that does not adhere to principles of integrity and ethical adequacy is a very dangerous pseudo-evaluation.
- b) Feasibility For any evaluation, it is necessary to consider the available resources (be they human, financial, or temporal) to ensure it can be conducted. If the conditions for feasibility are not met, then the evaluation may not even take place.
- c) Safety This aspect of evaluation is generally associated with evaluations of products in the pharmaceutical, food, automotive industries, children's toys, electrical equipment, among others. Naturally, every evaluation must be safe for all participants and stakeholders.
- d) Significance and Social Utility Evaluation is only worthwhile if it has utility. Otherwise, it is a waste of resources. Moreover, an evaluator should consider the possibility that their evaluation may have implications beyond its original scope.
- e) Equity This principle is linked to democratic societies. Equity is particularly important in performance evaluations of individuals, such as in assessing an employee within an organization or a student.

Beyond these values, the quality of an evaluation can be determined in two ways: through criteria or through lived experiences. In the first case, the goal is for quality to be independent of the evaluator. In criterion- or standards-based evaluation, there are four fundamental steps to consider: 1) defining criteria/standards, 2) defining indicators and descriptors for each criterion (or standard), 3) establishing a measurement process for these indicators or descriptors, and 4) analysing the data in light of the defined criteria/standards to determine the merit or value of what is being evaluated. In

the second case, through lived experiences, quality can be determined by describing, analysing, and discussing perceptions. Here, quality is socially constructed and recognized when seen or felt; quality thus becomes an experience (Fernandes, 2013).

It is important to remember that whether evaluation and quality discernment are based on criteria, standards or lived experiences, it remains a social practice and a human construct. Considering that the perception of quality is closely linked to each individual's lived experience (Stake & Schwandt, 2006), this human experience is also considered in the formulation (also human) of criteria and standards. Thus, even criteria- and standard-based evaluations are not fully free from the human and social aspects of evaluation.

Therefore, it is restated that in evaluating a specific reality or object, it must be considered that there is no absolute truth about what is being evaluated. Thus, the only way to ensure greater accuracy in evaluation is to anchor it in a solid theoretical framework, considering: a) the purposes and questions of the evaluation, b) the methodological perspectives adopted, procedures, and techniques to be emphasized; c) the role of the evaluators, d) the level of stakeholder participation, e) the definition of the target audiences, and f) the nature and dissemination of the evaluation report (Fernandes, 2010); all governed by integrity, ethical adequacy, feasibility, safety, significance, social utility, and equity (Stufflebeam & Coryn, 2014).

1.2. Educational Assessment

The terms "Evaluation" and "Assessment" are frequently used in educational contexts. In the United Kingdom, the common use of the term "assessment" is to refer to judgements of students' work, and "evaluation" to refer to judgements regarding courses or course delivery, or the process of making of such judgements (Taras, 2005). With shared foundational principles with evaluation, educational assessment constitutes itself as a process for gathering information that can be used to make decisions about students, teachers, curriculum, programmes, and schools, as well as other aspects of education systems and policy (NAE, 2021). This process occurs at different levels of education systems, ranging from the individual student assessment conducted by classroom teachers to national assessments and cross-national comparisons of student achievement carried out by national or international agencies (Kellaghan et al., 2003; Santiago et al., 2012).

The growing complexity of problems affecting education systems demands that assessments describe, analyse, and interpret the objects and phenomena of concern with greater depth, credibility, utility, and rigor (Fernandes, 2013). Therefore, a diversity of assessments and evaluations conducted

at different levels of education systems, with various dimensions, characteristics, participants and stakeholders is essential for better describing, understanding, and addressing these problems.

These different assessments that occur within education systems must be assembled with adhering principles that ensure that the elements are complementary and work together (Berman et al., 2020; Colardaci, 2002). Otherwise, they are completely independent assessments.

A collection of assessments does not entail a system any more than a pile of bricks constitutes a house (Coladarci, 2002: 773).

To hold assessments together where they work collectively to fulfil specific interpretive goals and purposes (Berman et al., 2020), an assessment system is one that is comprehensive, coherent, and involves the continuous collection of assessments (Coladarci, 2002; NRC, 2001b). Comprehensive means that different approaches to assessment are used to provide a variety of evidence to support decision-making. The use of multiple assessments and indicators enhances the validity and fairness of the inferences drawn, by giving students various ways and opportunities to demonstrate their competence. To support learning, the system must also be coherent, meaning that there is consistency in the conceptualization of student learning underlying the various assessments. In this way, results from external assessments will align with the more detailed understanding of learning that underlies classroom instruction and assessment. Additionally, an assessment system should be designed to be continuous, meaning that assessments should measure student progress over time to provide a clear picture of growth (Berman et al., 2020).

1.2.1. Student Assessment Purposes

The distinction between the Summative Assessment and Formative Assessment was first proposed by Scriven (1967) in the context of programme evaluation. For Scriven, summative evaluation provided information to judge the overall value of an educational programme, whereas the results of formative evaluation were targeted at facilitating programme improvement (Bennett, 2011). Later, in 1969, Benjamin Bloom made a similar distinction, using the same terminology of formative and summative evaluation with respect to students (see Bloom, 1969). Bloom suggested that the same distinction might be applied to the evaluation of student learning – what today we tend to call "assessment" – acknowledging the traditional role that tests played on summative purposes but noting that there was

another role for evaluation: the use of "formative evaluation" to provide feedback and correctives at each stage in the teaching-learning process (Dylan, 2006).

Since then, these two concepts have been further developed for educational settings and, nowadays, in education, formative assessment refers to assessments that provide information to students and teachers that is used to improve teaching and learning. These assessments are often informal and ongoing. Summative assessment refers to the cumulative assessments, usually occurring at the end of a unit or topic coverage, that intend to capture what a student has learned, or the quality of the learning, and judge performance against some standards. Still, data from summative assessments can also be used in a formative way (NRC, 2001a).

Characteristic	Formative Assessment	Summative Assessment
Purpose	To improve teaching and learning To diagnose student difficulties	Evaluation of learning outcomes Placement, promotion decisions
Formality	Usually, informal	Usually, formal
Timing of administration	Ongoing, before and during instruction	Cumulative, after instruction
Developers	Classroom teachers to test publishers	Classroom teachers to test publishers
Level of stakes	Low-stakes	High-stakes
Psychometric rigor	Low to high	Moderate to high
Types of questions asked	What is working? What needs to be improved? How can it be improved?	Does student understand the material? Is the student prepared for next level of activity?

Table 3. Characteristics of Formative and Summative Assessments

Source: Adapted from Dixson & Worrell, 2016:154

Despite this distinction that might be made on the purposes of assessments, Taras (2005) highlights that formative assessment is the same process as summative assessment. Though, since this distinction between summative and formative assessment emerged in educational settings, there has been some confusion between the process of assessment with its socio-educational functions. Hence, the focus on its many functions has compounded the perceived division between summative assessment, where summative assessment has been blamed for many

problems related to assessment in education systems and as an obstacle to the growth of formative assessment.

Taras (2005) also clarifies that the assessment process culminates in a summative assessment, which is a judgment that encompasses all the evidence collected up to a specific moment. Additionally, a summative assessment can serve multiple functions that do not interfere with the assessment process. Still, for an assessment process to be formative, it requires feedback which indicates the existence of a 'gap' between the actual level of the work being assessed and the required standard. It also requires an indication of how the work can be improved to reach the required standard and, therefore, teachers have a crucial role in providing that feedback (Gardner et al., 2010; Sadler, 1998).

Note as well that while the uses of assessments cannot be controlled or limited – i.e., even if we decide that an assessment should be created, undertaken and graded with a pre-determined function, there can be no guarantees of how it will be used, who will use it or when in the future the results of this assessment might be adopted with different functions than those intended – still, we can control the assessment processes parameters, i.e., the criteria, intended outcomes, and standards which form the basis of the assessment (Taras, 2012).

These two key concepts – summative assessment and formative assessment – are the basis for the current Portuguese Assessment System, that will be further developed in the Portuguese Evaluation and Assessment Framework' section of this chapter.

1.2.2. Assessment of Learning, for Learning and as Learning

An alternative way to conceptualize the purposes and uses of student assessment is by categorizing them into assessments *of* learning, assessments *for* learning and assessments *as* learning. Assessments of learning monitor knowledge and understanding as reflected in performance on tests, often in terms of measuring progress towards specific learning objectives. Assessments for learning enable teachers to use information about students' knowledge and skills to guide instruction and to offer feedback that helps students track and improve their learning. Finally, assessments as learning occurs when the assessment process itself not only tracks learning but also influences it (NAE, 2021).

Thirty years after Scriven's first definition of summative and formative evaluation, assessment for learning (or formative assessment) became the *core* of classroom assessment in many educational settings worldwide, especially after the seminal work of Black & Wiliam in 1998. With important contributions from scholars such as Perrenoud, Biggs, Hattie & Jaeger, Sadler, Sebatane and Dwyer, Black & Wiliam's work pushed the research field of educational assessment to a new direction, and assessment for learning has emerged as one of the most important purposes of assessment, offering

important research themes on feedback, self-assessment and formative assessment processes (Hopfenbeck, 2018).

Broadfoot et al. (2002) referrer that ten principles should be considered to guide classroom practice in assessment *for* learning:

- Assessment for learning should be part of effective planning of teaching and learning. Teachers should provide opportunities for both teachers and student collect and use information about progress towards learning goals. Learners should understand learning goals and the criteria that will be applied in assessing their work.
- 2. Assessment for learning should be focus on how students learn, i.e., the process of learning has to be in the mind of both learner and teacher.
- 3. Assessment for learning should be recognized as central to classroom practice. Tasks and questions prompt learners to demonstrate their knowledge and skills; and these assessments processes are an essential part of everyday classroom practice, involving both teachers and learners in reflection, dialogue and decision making.
- Assessment for learning should be regarded as a key professional skill for teachers. Teachers should be supported in developing these skills through initial and continuing professional development.
- 5. Assessment for learning should be sensitive and constructive. Teachers should be aware of the impact that comments, marks and grades can have on learners.
- Assessment for learning should take account of the importance of learners' motivation. Assessment that encourages learning by emphasising progress and achievement rather than failure.
- 7. Assessment for learning should promote a commitment to learning goals. For effective learning to take place, learners need to understand what is they are trying to achieve and, therefore, a shared understanding of the criteria by which they are assessed should be provided.
- 8. Assessment for learning provides to learners constructive guidance about how to improve.
- Assessment for learning develops learners' capacity for self-assessment to that they can become reflective and self-managing.
- 10. Assessment for learning should recognise the range of achievements of all learners, enhancing all learners' opportunities to learn, in all areas of educational activity.

However, Bennett (2011) recalls that formative assessment does not yet constitute a well-defined set of artifacts or practices. While research indicates that the general practices linked to formative assessment can support learning, the existing definitions are so broad that the effects are likely to vary significantly across different implementations and student populations. Taras (2012) also raises the question if assessment for learning is assessment or learning? A distinction between "summative assessment" and "formative assessment" only has any real meaning if formative assessment belongs to the learner. Learners are the ones who learn and who ultimately make all the decisions about learning, whether consciously or unconsciously. That is why theories of learning and assessment for learning should be developing more closely with each other (Baird et al., 2017), where assessment for learning should be situated within a social-constructivist theory of learning (Taras, 2012) or within the context of a theory of pedagogy (Black & Wiliam, 2018). Hence, if the main purpose of assessment and learning should be developed further and be at the forefront of high-stakes assessment and large-scale educational assessments. Still, assessment and learning theories seem to be fields apart, with scope for far greater connection (Baird et al., 2017).

1.2.3. Audiences

Assessment results can serve a wide range of audiences and users. For example, with the very same test score, students themselves can identify their learning gaps. Parents and caregivers might use the test score to understand how their children's opportunities and achievements compare to those of other students in the class or school. Teachers may use that test score to identify areas that require additional or differentiated instruction. On a broader scale, schools and states may use that test score to monitor overall student performance, identify and address systemic inequities, allocate resources, evaluate teacher effectiveness, and identify professional development needs (NAE, 2021). This range of interested parts (stakeholders) are often asking different questions about the same assessment, but answers to these questions do not always coincide with the interpretive uses for which the assessments were originally designed and validated (Berman et al., 2020). Therefore, the audiences and the purposes of assessments in Education are crucial in discussions of assessment theories, as they can shift our focus from the concrete results of assessments to their more unpredictable uses and impacts, obscuring its primary purpose (Taras, 2012).

These issues regarding student assessment will be further developed in the following chapters of this thesis, taking Portugal in the Covid-19 context as a case study.

1.3. The Portuguese Education System

1.3.1. Governance

In Portugal, education is a fundamental right established in the Constitution of 1976, following the April 25th Revolution of 1974. Education is organized according to the democratic principles of the Constitution, particularly Article 43 (freedom to teach and learn) and Articles 73 to 77 – which cover education, culture, and science (Article 73), teaching (Article 74), public, private, and cooperative education (Article 75), universities and access to higher education (Article 76), and democratic participation in education (Article 77).

Although the Portuguese Constitution dates back to 1976, the *"Lei de Bases do Sistema Educativo"* (LBSE) [Education Act], was only approved in 1986, 12 years after the revolution. With the Education Act (LBSE) approval, the guiding parameters for the structure and functioning of the education system were defined, outlining the principles to be followed by its administration and management structures at the central, regional, autonomous, local, and school establishment level (Formosinho & Machado, 2013; Rodrigues et al., 2016). With LBSE, it was also established compulsory school until the 9th grade and, later, in 2009, compulsory school was extended to the age of 18 years old with the approval of Law 85/2009.

School governance in Portugal is fairly centralised (Santiago et al., 2012), as Pre-primary education, basic and upper secondary education are a responsibility of the Ministry of Education. Therefore, the Ministry of Education is responsible for defining, coordinating, implementing and assessing national policy with regard to the education system, as well as articulating education policy with qualification and vocational training policies (Eurydice, 2024) through its multiple administrative structures, such as DGE - Direção-Geral da Educação [Directorate General for Education], DGAE -Direção Geral da Administração Escolar [Directorate General for School Administration], DGEEC -Direção-Geral de Estatísticas da Educação e Ciência [Directorate General for Education and Science Statistics], IGEC – Inspeção-Geral da Educação e Ciência [General Inspectorate of Education and Science], and IAVE – Instituto de Avaliação Educativa [Institute for Educational Assessment]. Note that, although governance is quite centralized, in recent decades, there has been some transfer of responsibilities to municipalities in areas such as curricular enrichment activities, management of school infrastructure, and management of non-teaching staff human resources (Santiago et al., 2012). Note as well that the Autonomous Regions of the Azores and Madeira, under Articles 225 to 234 of the Portuguese Constitution, can legislate to establish local educational policies through their Regional Governments and respective Regional Secretariats for Education.

Still concerning administrative and governance characteristics, most school clusters or nonclustered schools in Portugal have a regular administration directly dependent of the Ministry of Education, but some schools have an autonomy contract with the Ministry of Education for the development of their own educational project (see Decree-Law No. 75/2008). Moreover, the creation of *Territórios Educativos de Intervenção Prioritária* (TEIP) [Educational Territories of Priority Intervention] is a government initiative, currently implemented in 146 school clusters or non-clustered schools, to better support schools that are in economically and socially disadvantaged territories (DGE, 2023). In a new program phase (TEIP4), school clusters apply for support through an action plan in partnership with the municipalities.

The Portuguese school network is organized into school clusters and non-clustered schools, each with its own administration and management bodies. Most school clusters include both basic and secondary education (from 1st grade to 12th grade). A smaller number of school clusters consist only of basic education (from 1st to 9th grade). Non-clustered schools that offer only secondary education (from 10th to 12th grade) are the least common. School clusters, therefore, consist of several establishments across one or more educational levels, aiming to facilitate the transition between levels and cycles of teaching. They also aim to overcome the isolation of individual schools, prevent social exclusion, consolidate the pedagogical capacity of educational establishments, and ensure the rational use of resources (Santiago et al., 2012). School clusters and non-clustered schools have some degree of autonomy in areas such as pedagogy, managing teaching and non-teaching staff (see Decree-Law No. 75/2008; Martins & Albuquerque, 2019), as well as in the students' internal assessment process (see Decree-Law No. 55/2018). This latter topic will be discussed in detail in the following section 1.4. – Policies for Student Assessment in Portugal.

1.3.2. Compulsory Education System Structure

According to the Education Act (LBSE), the Portuguese education system is divided into three levels: pre-school education (ISCED 0), basic education (ISCED 1 & 2) and upper secondary education (ISCED 3). School is compulsory from the age of 6 until the age of 18 or up to the 12th grade. While the Portuguese education system is divided into pre-school education (from the age of 3 until the start of basic education), basic education (expected ages 6 to 15), and upper secondary education (expected ages 15 to 18), attending pre-school is optional, and attendance is only compulsory for basic and secondary education (Eurydice, 2024; Decree-law No. 85/2009; Decree-law No. 55/2018).

Within compulsory education, basic education is universal, and it has the same track for all students. However, in some schools there is artistic education, which adds complementary training in an artistic area to the general curriculum (specialised artistic courses) (see Decree-law No. 55/2018).

As depicted in Figure 2, basic education lasts nine years and it is divided into three sequential cycles:

- The first cycle (lower ISCED 1) corresponds to grades one to four.
- The second cycle (upper ISCED 1) corresponds to grades five and six.
- The third cycle (ISCED 2) corresponds to lower secondary education (grades seven to nine).

Expected Age (years)	6	7	8	9	10	11	12	13	14	15	16	17
Grade	1 st	2 nd	3 rd	4 th	5 th	6 th	7 th	8 th	9 th	10 th	11 th	12 th
General Courses												
	Single track								Multiple tracks			
Within School Clusters				Bas	ic educa	ition			Secondary general			
	1 st cycle 2 nd cycl				cycle	3 rd cycle			education			
	Ci	ursos ai	tísticos	especia	lizados [speciali	zed arti	stic cou	rses]			
Within School Clusters or Artistic schools Grade 5 – 12												
Cursos profissionais [vocational courses]												
Within School Clusters or Vocational schools									Grade	10 – 12		

Figure 2. Structure of the National Education System

Sources: Decree-law No. 85/2009; Decree-law No. 55/2018; adapted from Eurydice, 2024

Upper secondary education (ISCED 3) lasts for three years and corresponds to grades 10, 11 and 12 of upper secondary education. Upper secondary education aims to provide students with diverse training and learning, with a view to pursuing further studies and/or entering the labour market. Therefore, tracking only begins in secondary education (around 15 years old) with the assignment of students to different educational programs (Decree-law No. 85/2009; Decree-law No. 55/2018; Parveva et al., 2020).

School education in Portugal is predominantly funded by public funding (Parveva et al., 2020). Differentiation between types of school can occur due to differences in governance and funding (public or private sector); due to differences in the curriculum (e.g. schools offering different specialisations or educational programmes, such as *Escolas Artísticas* [artistic schools] and *Escolas Profissionais* [vocational schools]); or through structural features (different school types catering for different age groups or levels of education in parallel, such as school clusters from grade 1 to 9, school clusters from grade 1 to 12, or non-clustered secondary education school (grade 10 to 12).

1.3.3. Evaluation and Assessment Framework in Portugal

Assessment policies throughout the education system can have a profound influence on assessment practices, and enhancing assessment to improve learning requires consistent and coordinated actions at state, school and classroom levels (NRC, 2001a). In addressing the topic of evaluation and assessment framework in Portugal, it is therefore important to identify not only what is being assessed but also where the assessment is taking place (within the classroom, the school, or the educational system) and who the evaluated actors are (students, teachers, schools, or the educational system) (Pedrosa et al., 2022).

In the late 1960s, under the dictatorial regime, educational assessment in the Portuguese education system had primarily classificatory functions and was mainly used as a tool for selecting students through external exams (Fernandes, 2009; 2014). Fifty years later, amid a revolution in 1974, the Portuguese education system has undergone a profound transformation in its multiple dimensions, a change that the book 40 Anos de Políticas de Educação em Portugal – Volume I & II [40 Years of Education Policies in Portugal], with texts from various educational researchers and organized by Maria de Lurdes Rodrigues portray very well (See Rodrigues, 2014a, 2014b). Additionally, the advancement of educational assessment in Portugal has been remarkable, and Portugal has come far in developing the foundations of a framework for evaluation and assessment (Santiago et al., 2012).

With the aim of improving educational assessment and, consequently, enhancing student learning, educational assessment has evolved significantly in its internal and external processes. This includes the introduction of low stake assessments of students' learning (such as *Provas de Aferição*), students' international large-scale assessments (such as PISA), teachers appraisal, school evaluation, and educational programs promoted by local and/or central government evaluations. Still, the creation of an assessment system requires significant thought, effort, time, and resources, and it is not established quickly or all at once: it evolves (Coladarci 2002). Nowadays, although Portugal has already established a relatively elaborate system of accountability (Parveva et al., 2020), policy initiatives in evaluation and assessment for many years, have emphasised accountability over improvement. Hence, the Portuguese assessment system still faces many challenges as there is no integrated evaluation and assessment framework or, at least, it is not perceived as a coherent whole and it does not visibly connect all the different components (Santiago et al., 2012).

Portugal does not have a single evaluation and assessment framework that was designed as a whole, but instead has a series of components operating at different levels that have developed relatively independently of each other over time (Santiago et al., 2012: 26). Evaluation and assessment operate at four key levels: System, School, Teachers and Student:

- Considering the dispositions of article 52 of Education Act (LBSE), system evaluation draws mostly on the evaluation of schools, complemented with external student assessment, a wide variety of indicators on education based on data collected from schools on a standardised format, on international benchmarks of students' performance such as PISA, TIMMS and PIRLS, and on studies of the impact of policy initiatives (Santiago et al., 2012).
- 2. At a school level, besides self-evaluations, schools have external inspections, which are a responsibility of IGEC. The 1st cycle of external assessment started in 2006 until 2011. The 2nd cycle started in 2011 until 2017 and, currently, schools are facing the 3rd cycle of external evaluations that started in 2018. According to IGEC (2024), external inspections aim to:
 - a) Promote the quality of teaching, learning, and the inclusion of all students;
 - b) Identify strengths and priority areas of intervention to improve the planning, management, and educational action of schools;
 - c) Assess the effectiveness of schools' self-evaluation practices;
 - d) Foster a culture of participation within the educational community;
 - e) Contribute to the public understanding of the quality of schools' work;
 - f) Produce information to support decision-making in the development of educational policies.
- 3. Teachers appraisal formally began in 2007 with the approval of Decree-Law No. 15/2007. It then underwent several revisions and changes in the following years and, currently, teacher appraisal is regulated by Regulatory Decree No. 26/2012. Aiming to improve the quality of educational services and student learning, as well as to enhance and support the personal and professional growth of teachers, it focuses on the following dimensions:
 - a) Scientific and pedagogical dimensions;
 - b) Participation in the school and relationship with the community;
 - c) Continuous training and professional development.
- 4. Student performance is assessed by a wide range of instruments, ranging from ongoing daily formative assessment in the classroom to national and international standardised tests.

Although there would much to explore, report, and study within this complex evaluation and assessment framework with its multiple relationships within the education system, the scope of this thesis focuses on student assessment during the Covid-19 pandemic. Therefore, only national student assessment policy is further developed in the next section.

1.4. Policies for Student Assessment in Compulsory Education

In the tumultuous years following the 1974 revolution, political instability led to the rise and fall of eight Ministers of Education, none of whom completed a full term. During this period, student assessment was not a priority in educational policies. Despite some guidelines suggesting that assessment should be more focused on learning, teachers were largely left on their own with their assessments, navigating contradictions between the "Estado Novo pedagogy" and the challenges of a "revolutionary school serving a democratic and socialist society" (Fernandes, 2014).

Later, since the approval of the Education Act (LBSE) in 1986, student assessment in compulsory education has undergone several reforms, reflecting that an assessment system is not established quickly or all at once (Coladarci, 2002). Table 4 highlights key developments in student assessment within compulsory education, beginning with the Education Act (LBSE) in 1986, followed by the approval of Normative Dispatch No. 98A/92 in 1992, which places formative assessment at the core of compulsory education, and extending up to the onset of the pandemic in 2020.

Date	Legislation	Observations
1986	Education Act (LBSE): Law No. 46/86	Establishes the general framework of the educational system. Basic education becomes compulsory for children and youth between the ages of 6 and 15 years old.
1992	Normative Dispatch No. 98-A/92	Approves the student assessment system for basic education.
1994	Normative Dispatch No. 644-A/94	Amends the Normative Dispatch No. 98-A/92, introducing <i>Provas Globais</i> in basic education.
2000	Normative Dispatch No. 5437/2000	Determines the conditions of <i>Provas de Aferição</i> (PA) implementation at the end of the three cycles of basic education.
2001	Decree-Law No. 6/2001	Approves the curricular reorganization of basic education. Defines diagnostic, formative and summative assessment.
2002	Decree-Law No. 209/2002	Defines Summative internal and external assessment.
2009	Law No. 85/2009	Establishes the compulsory education system for children and youth between the ages of 6 and 18 years old.
2012	 Decree-Law No. 139/2012	Establishes principles for curricula organization and management, and student assessment in basic and secondary education.
2016	Normative Dispatch No. 1-F/2016	Regulates assessment and certification of learning.
2017	Dispatch No. 6478/2017	Introduces the <i>Perfil dos Alunos à Saída da Escolaridade</i> <i>Obrigatória</i> [Students Profile by the End of Compulsory Education]
2018	Decree-Law No. 55/2018	Establishes the curriculum for basic and secondary education and the guiding principles for the assessment of learning.
2018	Decree-Law No. 54/2018	Establishes the legal framework for inclusive education.

Table 4. Synthesis of Portuguese Legislation

Sources: Fernandes, 2014; adapted from Pedrosa et al., 2022:61

Note that the expansion of compulsory schooling in 1986 (up to the age of 15) and in 2009 (up to the age of 18) introduced significant challenges related to student assessment policy, practices, purposes, and functions. Hence, while coercive measures ensured that all children had access to school, these two periods of compulsory schooling expansion in the democratic regime are also characterized by high levels of grade repetition (Justino, 2014) with persistent educational inequalities (Martins, 2016).

Note as well that pre-school is not part of compulsory education in Portugal. Additionally, preschool has specific curricular and assessment guidelines (OCEPE) designed for Early Childhood Education (see DGE, 2023). Within the scope of this thesis, in the following sub-sections, only the student assessment framework for compulsory education in 2020 will be analysed, the year in which the Covid-19 pandemic was declared. This analysis includes the existing legal framework, as well as two assessment projects with a strong impact on student assessment: the Training, Supervision and Research in Classroom Assessment (MAIA) project, aimed at enhancing teachers' assessment literacy and practices through professional development (Fernandes, 2021; MAIA project, 2023); and the Dematerialization of External Assessment (DAVE) project, which seeks to integrate and dematerialize all procedures related to external assessment processes (IAVE, 2022)

1.4.1. The PASEO

The PASEO – *Perfil dos Alunos à Saída da Escolaridade Obrigatória* [Students' Profile by the End of Compulsory Education] is an innovative and remarkable declaration that aims to provide a unique reference document to ensure the coherence of the entire education system and, thereby, give meaning to compulsory schooling (PASEO, 2017). After extensive public discussion, the PASEO was approved in 2017 by Dispatch No. 6478/2017. As a result, the PASEO serves as a reference for organizing the entire educational system and contributes to the convergence and coordination of decisions related to various dimensions of curriculum development by educational managers and stakeholders, both at the policy-making level and within educational institutions. Its goal is to support curriculum organization and management, as well as to define strategies, methodologies, and pedagogical-didactic procedures to be used in teaching practice (DGE, 2024).

The document assumes a necessarily broad, transversal, and recursive nature. The broadness of PASEO respects the inclusive and diverse character of schools and students. Transversality is based on the assumption that each curriculum area contributes to the development of all competence areas. The recursive nature of the document lies in the possibility that, in each year of schooling, its content and purposes are continually invoked (PASEO, 2017).

The *Students' Profile* document is structured in Vision, Principles, Values and Competence Areas. The Vision within the PASEO embodies designs that complement each other, aimed at both individual qualification and democratic citizenship. To students at the end of compulsory education are expected to be citizens who:

- Develop multiple literacies, enabling them to critically analyse and question reality, evaluate and select information, formulate hypotheses, and make informed decisions in their daily lives;
- Are free, autonomous, and responsible, self-aware, and aware of the world around them;
- Are able to cope with the transformation and uncertainty of a rapidly changing world;
- Acknowledge the importance and challenges presented by the Arts, Humanities, Science, and Technology for the social, cultural, economic, and environmental sustainability of Portugal and the world;
- Are autonomous and able to use a range of developed skills, including critical thinking, creativity, collaborative working skills, and communication skills;
- Are able to continue lifelong learning as a decisive factor in their personal development and social participation;
- Understand and respect the fundamental principles of democratic society and the rights, guarantees, and freedoms on which it is based;
- Value respect for human dignity, the exercise of full citizenship, solidarity with others, cultural diversity, and democratic debate;
- Reject all forms of discrimination and social exclusion.

To this end, the PASEO is based in eight principles:

- 1. Humanistic-based profile School empowers young people to build a fairer society, centred on the individual, human dignity, and the preservation of the world as a common good.
- 2. Knowledge Knowledge is in the centre of the education process, enabling students to understand, make decisions and act upon the world's natural and social realities.
- Learning Educational action deliberately promotes the development of learning skills, the basis for education and lifelong learning.
- Inclusion Compulsory schooling is for everyone, fostering equity and democracy by bringing together a diverse range of students, not only from socioeconomic and cultural backgrounds but also from cognitive and motivational perspectives.
- 5. Coherence and flexibility Ensuring access to learning and student participation in their training process requires coherent and flexible educational action.

- Adaptability and audacity Educating in the 21st century demands one's awareness to be able to adapt to new contexts and new structures.
- Sustainability School contributes to raise students' awareness of sustainability, one of the greatest challenges in the contemporary world.
- 8. Stability The PASEO enables coping with evolution in every area of knowledge and get stability so that the system may adjust and produce effects.

At the *core* of this framework are five values:

- Responsibility and integrity Self-respect and respecting others; knowing how to act ethically and being aware of their own actions in the light of the common good.
- 2. Excellence and demand Aspiring to the achievement of a well done work, of rigour and of overcoming; being aware of themselves and others with sensitivity and solidarity with others.
- 3. Curiosity, reflexion, innovation Willing to learn more; developing reflective, critical and creative thinking; striving for new solutions and applications.
- 4. Citizenship and participation Demonstrating respect for human and cultural diversity and acting in accordance with human rights principles.
- 5. Freedom Show personal autonomy centred in human rights, rights of democracy, citizenship, equity, in mutual respect and in the free choice of the common good.

Finally, there is a set of competences to be valued. Still according to PASSEO (2017), competences are complex combinations of knowledge, skills and attitudes, and they are vital for the students' profile, as well as for compulsory schooling. Competences to be considered are:

- a) Languages and texts;
- b) Information and communication;
- c) Reasoning and problem solving;
- d) Critical and creative thinking;
- e) Interpersonal relations;
- f) Personal development and autonomy;
- g) Well-being, health and environment;
- h) Aesthetic and artistic sensitivity awareness;
- i) Scientific, technical and technological knowledge;
- j) Body awareness and mastery.





Source: Adapted from PASEO, 2017:10

The PASEO' principles, values, and competence areas entail changes in pedagogical and didactic practices to align the overall educational action with the PASEO purposes and vision. The educational process is, therefore, understood as a specialized formative action aimed at achieving learning outcomes, i.e., it is about finding the best way and the most effective resources for students to learn (PASEO, 2017). Such a structural document naturally has strong implications for student assessment purposes and practices, where the centrality of assessment *for* learning gains renewed emphasis.

1.4.2. The Decree-Law No. 55/2018

Aiming to align educational policy with PASEO, Decree-Law No. 55/2018 was produced in 2018. This Decree-Law redefines the primary and secondary education curriculum, guiding principles, and student assessment processes. The latter are detailed in Articles 22 to 32, in Section III of Chapter II of this Decree-Law.

1.4.2.1. Assessment Purposes

According to Article 22 of Decree-law No. 55/2018, student assessment should be grounded in a formative dimension, as an integral part of teaching and learning processes. Student assessment should also guide students' educational paths and certify the knowledge, skills, and attitudes developed according to the PASEO. To this end, it should involve the use of diverse and appropriate assessment methods. Both internal and external assessments processes should complement each other to inform and support pedagogical interventions, readjusting strategies that lead to the improvement of learning quality. Hence, they should assess the achievement of the objectives defined in the curriculum and certify learning.

1.4.2.2. Internal Assessment

Articles 23, 24 and 27 define internal assessment as formative or summative.

Formative assessment assumes a continuous and systematic character, serving learning processes, employing a variety of procedures, techniques, and instruments for gathering information, suitable for the diversity of learning, recipients, and circumstances in which they occur. Summative assessment translates into a comprehensive judgment of the learning achieved by students, with the aim of grading and certification (Decree-Law No. 55/2018).

Furthermore, Articles 24 and 27 refer that formative assessment should be the primary mode of assessment as it provides privileged and systematic information across various curricular areas. It should, therefore, support learning by involving students in the self-regulation process and be integrated with informational tools aimed at parents and guardians (Decree-law No. 55/2018).

1.4.2.3. External Assessment

External assessment processes are defined with Articles 23 and 25 of Decree-Law No. 55/2018. The process is both centralised and totally controlled by the Ministry of Education, with schools and teachers cooperating in its administration and grading, under the control of the National Examination Board and other Ministry of Education departments (Fernandes, 2009). External assessment tests are

a responsibility of *Instituto de Avaliação Educativa* (IAVE) [National Institute of Educational Assessment] and they are compulsory for all students in the Portuguese education system.

As Figure 4 shows, in the middle of each cycle of basic education, in the 2nd, 5th, and 8th grades, there are *Provas de Aferição* (PA). At the end of basic education, in the 9th grade, there are *Provas Finais de Ciclo* (PFC). At the end of secondary education, in the 11th and 12th grades, there are *Exames Nacionais* (EN). The conclusion of secondary education on the artistic or professional path does not imply performing EN. There is, in replacement, *Provas de Aptidão Artística and Provas de Aptidação Profissional*, which are aptitude artistic or vocational tests set by each school (Decree-law No. 17/2016; Decree-law No. 55/2018).

Level/Cycle	Basic education									Cocondamy advection		
		1 st c	cycle		2 nd (cycle		3 rd cycle	1	Secondary education		
Grade	1 st	2 nd	3 rd	4 th	5 th	6 th	7 th	8 th	9 th	10 th	11 th	12 th
External tests	-	PA	-	-	РА	-	-	PA	PFC	-	EN	EN
Notes. PA – Provas de Aferição; PFC – Provas Finais de Ciclo; EN – Exames Nacionais												

Sources: Decree-law No. 85/2009; Decree-law No. 55/2018

PA, in the 2nd, 5th, and 8th grades, are national assessment tests to assess pupils' learning and they do not have weighting on students' grading. Some of these PA are practical tests with standardized procedures, such as PA in Music Education, Sports, and Arts; while other PA are written tests, such as in History and Natural Sciences. These tests aim to monitor the curriculum development, in different areas, providing regular information to the education system. They also aim to provide detailed information about student's performance to the school, teachers, guardians, and the students themselves, before the end of a cycle of studies, to promote a timely pedagogical intervention, considering the difficulties identified for each student in each subject area.

By the end of basic and secondary education, PFC and EN have a weight of 30% on subject areas with external assessment. Portuguese and Mathematics PFC are compulsory at the end of the 9th grade, and they aim to evaluate students' performance, certifying basic education' conclusion, creating the possibility to pursue different school paths in secondary education.

Compulsory EN, at the end of the 11th (2 exams) and 12th grades (2 exams), aims to evaluate students' performance and certify the conclusion of secondary education. In addition, they may also be considered for access to higher education purposes (Decree-law No. 17/2016; Decree-law No. 55/2018).

Note that Point 3 of Article 23 of Decree-law No. 55/2018 allows for tests and examinations to be conducted in electronic format, paving the way for the introduction of computer-based large-scale testing. This topic will be further developed and explored in Chapter 7 of this thesis.

Note as well that, since the approval of the Education Act (LBSE) in 1986, external assessments have undergone several reforms. For example, secondary education external examinations were only reintroduced in 1995/1996 after several years of intense and passionate national discussions, as well as ideological and pedagogical disputes and controversies (Fernandes, 2009). Additionally, PFC for Mathematics and Portuguese in grade 9 were introduced in 2004/2005 school year. More recently, during the course of this investigation and thesis writing, the weight and number of external assessments to be conducted in secondary education has also change and will undergo minor changes starting in 2024 (see Decree-Law No. 62/2023). Additionally, with the shift in the political cycle in 2024, all PA will be abandoned and replaced with *Provas de Monitorização das Aprendizagens* (ModA) for the 4th and 6th grades (See the announcement from the Council of Ministers dated July 18, 2024 – Council of Ministers Announcement, 2024).

1.4.2.4. Grade Transition

Compared to 2009/2010, fewer European education systems allow grade repetition. The number of education systems where grade progression is automatic has increased from four to six in primary education, and from two to four in lower secondary education (Parveva et al., 2020). Despite the recent writing of Decree-Law No. 55/2018, in Portugal grade repetition is still allowed, and it is regulated in Articles 27, 29, and 30 of Decree-Law No. 55/2018. According to these Articles, the evolution of the educational process for students in general basic education should follow a cycle logic, with students progressing to the next cycle upon meeting the learning objectives defined for each teaching cycle.

If a student does not achieve the learning objectives defined for a non-terminal year of a cycle, it may exceptionally be decided to retain the student in the same year of schooling. However, this does not apply to the 1st year of schooling. In the event of grade repetition, the curricular planning instrument for the class the student will join in the subsequent school year must include multi-level measures for curriculum access, defining appropriate teaching and learning strategies, as well as educational resources to fully support the development of the learning objectives for the student.

However, although Portuguese legislation specifies that in basic education grade repetition should only occur in the middle of a cycle under exceptional circumstances, Cipriano & Martins (2021) demonstrated that Portuguese teachers hold their own conceptions and beliefs about assessment purposes and practices, superseding official recommendations and legislation. As such, in general, there is a traditional approach to the organisation of classrooms in Portugal, and assessment for learning is not systematically used in Portuguese schools (Santiago et al., 2012). Research has also been showing that in many Portuguese classrooms internal assessments are more oriented and organised to classify and to rank pupils' achievements rather than to help them to learn. Assessment for grading, selecting and certifying pupils continues to be the predominant aim (Fernandes, 2009). Therefore, despite a decrease in grade repetition rates over time, grade repetition remains a common practice across all grade levels, with rates consistently higher at the beginning of cycles in basic education (i.e., in grades 2, 5, and 7), contrary to the specifications of Decree-Law No. 55/2018, as shown in Table 5. These trends are also illustrated in Figure 7 of Chapter 2 (p. 60), which contains institutional data from DGEEC (2023), showing grade repetition and dropout rates per academic year and grade level from 2015 to 2022.

	2015	2016	2017	2018	2019	2020	2021	2022
Grade 1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Grade 2	9.3	8.9	7.4	6.6	4.9	3.2	4.2	3.7
Grade 3	4.1	3.0	2.2	2.2	1.5	1.0	2.0	1.6
Grade 4	2.5	2.5	2.1	2.0	1.7	1.4	2.1	2.1
Grade 5	8.5	6.8	6.1	5.6	4.0	2.5	3.0	3.0
Grade 6	8.7	6.7	5.6	5.0	3.7	2.3	3.5	3.3
Grade 7	15.4	12.6	11.4	9.8	7.0	4.2	5.7	5.7
Grade 8	10.3	8.0	6.7	6.8	4.7	2.7	4.2	4.3
Grade 9	10.7	9.0	7.0	6.5	5.5	2.2	2.8	3.7
Grade 10	12.5	12.9	12.6	11.1	10.7	7.3	8.4	9.3
Grade 11	8.6	7.1	7.1	7.0	6.7	3.3	3.7	4.3
Grade 12	29.9	28.2	26.2	24.5	22.6	15.4	12.9	12.2

Table 5. Grade repetition and dropout rates, by academic year and grade level (%)

Source: DGEEC, 2023 through PORDATA

1.4.3. The Decree-Law No. 54/2018

In addition to Decree-Law No. 55/2018, on a commitment with UNESCO to reduce and eliminate exclusion within and from education, the Portuguese government approved, also in 2018, the Decree-Law No. 54/2018. This decree-law establishes the principles and norms that guarantee inclusion, as a process that aims to respond to the diversity of pupils needs. It also identifies measures to support learning and inclusion, as well as specific resources to be mobilized to meet the educational needs of each and every child and young person along their school path. The measures to support learning and inclusion are organized into three levels of intervention: universal, selective, and additional.

- Universal measures correspond to the educational responses that the school has available for all students, with the goal of promoting participation and improving learning outcomes.
- Selective measures aim to address the learning support needs that are not met by the implementation of universal measures.
- Additional measures aim to address significant and persistent difficulties in communication, interaction, cognition, or learning that require specialized resources for learning support and inclusion.

In each school, a multidisciplinary team is established to support inclusive education. Special education teachers, within the scope of their specialty, should support other teachers, in a collaborative and co-responsibility way, in the definition of pedagogical differentiation strategies, in learning reinforcement and in the identification of multiple means of motivation, representation and expression. Therefore, special education teachers are compelled to mobilize and define pedagogical differentiation strategies to meet each student educational needs along their school path, considering their circumstances. Note that, still according to Decree-Law No. 54/2018, schools must ensure that all pupils have the right to participate in assessment processes, and those adaptations to internal assessment processes are a school responsibility, with the obligation to make public the results of students' assessment. To do so, teachers should adjust their internal assessment processes to their students' needs, with changes such as:

- a) Diversification of instruments for the collection of information;
- b) Test/exam sheets in accessible formats;
- c) Interpretation in Portuguese Sign Language;
- d) Use of assistive products/devices;
- e) Extra time for tests;
- f) Answers' transcription;

- g) Reading of the test/exam sheets;
- h) Use of a separate room;
- i) Supervised breaks;
- j) Colour identification code in the test/exam sheets.

Regarding external assessments, schools should determine and communicate any necessary adaptations to the Ministry of Education, which must be included in the student's record. These adaptations may include:

- a) The use of assistive products;
- b) Leaving the room during the test/exam;
- c) Adapting the space or materials;
- d) The presence of a Portuguese sign language interpreter;
- e) Consulting a Portuguese language dictionary;
- f) Taking adapted tests.

In secondary education, it may also be required:

- a) Taking the Portuguese as a Second Language (PL2) exam;
- b) Being accompanied by a teacher;
- c) The use of support tools for applying grading criteria for students with dyslexia, as provided in the Regulations for external assessments;
- d) The use of additional time.

1.4.4. The MAIA Project

With the approval of PASEO, Decree-Law 55/2018, and Decree-Law 54/2018, it is possible to observe that the political rhetoric in recent years has been prioritizing the improvement of student learning. In this sense, the need to improve teachers pedagogical practices has also emerged, particularly in instruction and assessment. It is in this context that it was created the continuous teacher training project for non-higher education, titled Training, Supervision, and Research in Classroom Assessment (MAIA Project) (Fernandes, 2021).

The MAIA Project is a nationwide initiative with voluntary participation that began in September 2019, a few months before the Covid-19 lockdowns. Since the beginning of the MAIA Project, there was a systematic emphasis on the idea that student assessment is fundamentally a pedagogical

process. The emphasis on this idea was a critical issue in fulfilling the broader purpose of the project: to contribute to the improvement of learning for all students (Fernandes et al., 2021).

The MAIA Project is, therefore, a multidimensional project with a high degree of complexity, considering, among other aspects, the diversity of stakeholders, its national scope, and especially the assessment culture that is deeply rooted in the Portuguese education system (Fernandes et al., 2021). According to Fernandes (2021), the project was designed and developed based on six dimensions: 1) the theoretical and foundational dimension, 2) the conceptual dimension, 3) the training dimension, 4) the follow-up dimension, 5) the monitoring dimension, and 6) the research dimension:

- The theoretical and foundational dimension aims to analyse and discuss the epistemological and ontological foundations of different perspectives on evaluation in general, in order to gain a well-grounded understanding of various conceptions.
- The conceptual dimension, intrinsically linked to the previous dimension, aims to analyse and discuss the so-called foundational concepts of assessment, such as formative assessment, summative assessment, feedback, and assessment criteria.
- 3. The training dimension, crucial for the development of the intervention projects, based on principles of active teacher training, activity theory, and communities of practice, aims to develop teachers as reflective professionals who are capable of analysing their actions in order to improve them.
- 4. The follow-up dimension aims to define the procedures that should be followed to support participants in developing the actions necessary to achieve their project's objectives.
- 5. The monitoring dimension primarily involved discussing progress updates regarding the development of the training processes. Monitoring aims to function as a process and an opportunity for regulating and self-regulating the training processes.
- 6. Finally, in the research dimension, methods and procedures for data collection are defined to describe, analyse, and interpret the realities inherent to the MAIA project development.

The project is also organized into three Moments:

- 1. The first moment, which took place in September and October 2019, involved a 30-hour training program for instructors.
- 2. The second moment focuses on developing Intervention Projects with teachers. In this moment, instructors first organize and develop training processes to foster collaboration among teachers and to implement essential guidelines for pedagogical assessment, learning, and teaching. Teachers should then engage in in-depth discussions on theoretical and

conceptual issues to create their Intervention Projects. Although initially scheduled from mid-March to June 2020, the initial plan had to be adjusted due to school closures from the Covid-19 pandemic. Despite this, the MAIA project continued with distance training.

3. The third and final moment is designed for participants, along with their instructors, to create real opportunities to implement these Intervention Projects. This phase encompasses all actions and procedures needed to investigate various aspects related to the training dynamics throughout the project's development (Fernandes, 2021).

Some of the implications of MAIA project will be further developed in Chapter 6 of this thesis.

1.4.5. The DAVE Project

When Decree-Law 55/2018 was written in 2018, it was admitted for the first time in Portuguese legislation that external assessments could be performed in electronic format:

The tests and exams referred to in the previous point can be conducted in electronic format (Point 3, Article 23 of Decree-Law 55/2018).

Considering the technological advancements over the past decades, according to IAVE (2022), major international student assessments, such as PISA, TIMSS, and PIRLS, have long been conducted electronically, resulting in notable gains in validity, reliability, and comparability of results, as well as increased efficiency in the administration of assessments across participating countries. Therefore, the digital transformation of external assessments is regarded as a true priority, as it offers numerous technical benefits and opportunities, allowing for significant improvements in the development, implementation, and scoring of external assessments (IAVE, 2022).

DAVE aims to integrate and dematerialize all the procedures inherent to a national external assessment of learning process, from the organizational and logistical point of view, until the processes of test preparation, application, and marking. The implementation of DAVE is organized into five major projects, complemented by smaller-scale projects:
- GAEBS project: Involves the development of an application designed to integrate all the software applications that support the organizational and logistical process of administering external assessment tests;
- 2. Digital transition project: A consultancy project with the support of an international entity with established expertise in electronic large-scale assessments;
- 3. Instruments development project: A project that enables the development of external assessment instruments in collaboration with the National Cybersecurity Commission;
- 4. Application and administration project: A project that enables students to take external assessment tests in electronic format, both online and offline;
- 5. Supervision of scoring project: A project that facilitates the electronic scoring of external assessments, allowing for greater speed in the process and increased reliability between scorers.

To implement DAVE with the greatest security and contribution of all actors involved in the process, DAVE has been tested by IAVE since 2018, and it is currently being gradually developed until its full implementation in 2025 (Eurydice, 2024; IAVE, 2022).

Year	Provas de Aferição (PA)	Provas Finais de Ciclo (PFC)	Exames Nacionais (EN)
2022	Sample (pilot)		
2023	Universal	Sample (pilot)	
2024	Universal	Universal	Sample (pilot)
2025	Universal	Universal	Universal

Table 6. DAVE implementation plan

Source: Adapted from IAVE, 2022:2

Note that amid several controversies, the universal implementation of DAVE first occurred in 2023 with PA. In 2024, there was an intention to implement PFC universally in a digital format, along with PA (see Normative Dispatch No. 4/2024). However, due to a governmental shift in 2024, this implementation process was postponed, resulting in changes to the Table 6 plan. Additionally, the full paper-free system was also altered. In 2025, the Portuguese PFC will be entirely paper-free, while the Mathematics PFC will involve a hybrid application process. Note as well that in basic education, PA will be cancelled and ModA will be introduced. Hence, in secondary education, exams will remain paper-

based, but the marking process will be electronic (While no legal framework is yet available; see Council of Ministers Announcement, 2024). This implementation process will be further developed in Chapter 7 of this thesis.

1.5. Higher Education Access Regime

Although compulsory education should be regarded as a terminal stage of students' pathways as expressed in the PASEO, for students wishing to enrol in a higher education institution afterwards, the conclusion of compulsory education has significant implications for access to higher education and, consequently, affects the functioning of secondary education itself (Lourtie, 2020). Therefore, this section aims to provide an overview of the national policies related to the secondary education conclusion and access to higher education.

The legal framework for access to higher education is regulated by Decree-Law No. 296-A/98. The general access regime is the main pathway to access higher education, through which most applications are made for courses offered at public higher education institutions. However, the higher education admission system has undergone several reforms since the approval of Decree-Law No. 296-A/98 in 1998. These reforms primarily involve amendments to the general access regime, including changes such as the creation of special quotas for specific groups (e.g., candidates from the Autonomous Regions of Azores and Madeira, Portuguese emigrants, or candidates with physical or sensory disabilities) and the creation of special admission regimes (e.g., the Regime for diplomatic missions abroad, the Regime for students from Portuguese-speaking African countries, or the regime for students older than 23) (CRUP, 2024).

Exams taken by students to complete compulsory education can be considered for access to higher education in the general access regime. Therefore, according to DGES (2024), all students who have completed compulsory education and taken the national exams (EN) can apply for admission to a specific course within a specific higher education institution. Note that students who wish to apply for admission to a public higher education institution must take the EN corresponding to the entry requirements for the courses to which they are applying. The entry exams required for each course are set by each higher education institution, and generally cannot exceed two exams.

Note as well that admissions to each higher education institution and course is subject to quantitative limitations based on the number of places allocated annually. The number of places for each course at each public higher education institution is determined annually by the institutions themselves, taking into account their available resources, and according to government policy for higher education (DGES, 2024).

Hence, in the application to public higher education through the general access regime, each student can apply to a maximum of six pairs institution/course, in each of the three phases of the general access regime. The ranking of candidates for each course at each higher education institution is done based on a descending order of an application score, calculated using the following criteria:

- Final secondary education grade, with a weight of no less than 50%;
- Entrance exam scores with a weight of no less than 35%;
- Classification of ranking pre-requisites (when required) with a weight of no more than 15%.

For the purposes of access to higher education, the final grade of the secondary education course is calculated to the decimal point, without rounding, and converted to a scale of 0 to 200 (DGES, 2024).

Note that a same EN is counted twice in the general access regime, as it is used to determine both the final secondary education grade and the entrance exam score. As each university and polytechnic school can choose the weight they assign to the final secondary education grade and the entrance exam score, the final secondary education grade can have a weight up to 65% for access to higher education, while exams can have a weight up to 50%. In cases where, for example, universities use one exam score with a weight of 50% for access to higher education, considering that this very same exam score was already used for the conclusion of secondary education in subject areas with a national compulsory exam, the weight of this single exam can, in fact, exceed 50% for access to higher education, as it is counted twice.

1.6. Chapter Final Consideration

Within Chapter 1 of this thesis, the goal was to underline the complexities that exist in the field of educational evaluation and assessment, and to provide an overview of the Portuguese educational system and its assessment framework, focusing on the student assessment level. This overview emphasized recent political efforts to place learning at the centre of the Portuguese education system through the production of key documents such as the PASEO, Decree-Law 55/2018, and Decree-Law 54/2018. These three documents therefore renew the importance and centrality that assessment *for* learning should have within the Portuguese education system.

The reinforcement of assessment *for* learning through legal diplomas was also accompanied by two important projects with major implications for student assessment: the MAIA project, which aims to enhance teachers' assessment literacy and practices, thereby improving student learning; And the DAVE project, which focuses on the digital transition of the national large-scale testing system.

Note that although the legal framework for student assessment in the Portuguese compulsory education system is primarily formative, by the end of compulsory education, the use of summative assessments has a significant impact on students' ability to access higher education. This additional use has repercussions on the functioning of secondary education (Lourtie, 2020), with unpredictable effects and impacts that obscure assessments' primary purpose (Taras, 2012).

It is within this evolving complex assessment framework that the Covid-19 pandemic emerged in 2020. Such a disruptive context had major implications on student assessment, which its immediate impacts are studied in the following Chapter 2 of this thesis.

CHAPTER 2

Student Assessment During the Covid-19 Pandemic

Chapter Summary:

The state of emergency provoked by the Covid-19 pandemic, characterized by uncertainty regarding the risks to public health, imposed to the Portuguese government, schools, and teachers a dynamic management of this scenario of crisis. In this context of great unpredictability, with the physical space of schools closed and with the establishment of emergency remote teaching, the traditional challenges of assessing learning have become even more complex. Faced with this extensive disruption, the Portuguese government decided to produce extraordinary legislation to regulate both external and internal assessment processes. Considering the challenges and solutions found by the Portuguese government, schools, and teachers, this whole period of confinement sparked, not only in Portugal but also around the world, a forced reflection on the practices, purposes, and uses of assessments, questioning the centrality of some assessment instruments such as tests or exams. With this thesis chapter, through the analysis of several studies and statistical indicators, it is intended to highlight some immediate impacts of the pandemic on assessment practices in Portugal. It is also intended to address some implications of the policies produced by the Portuguese government during this initial phase of the pandemic on the investigation carried out in this thesis are addressed.

2.1. The Covid-19 Pandemic Context

In December 2019, the outbreak of a new and unknown disease was first reported in Wuhan, China, which, a few months later, provoked a global public health crisis. In Portugal, the book *Um Olhar Sociológico sobre a Crise Covid-19 em Livro* [A Sociological Glance at the Covid-19 Crisis in a Book] (Carmo et al., 2020) provides an insightful perspective on its multiple impacts on Portuguese society.

The timeline presented in the following section of this chapter was created in 2020 as the events unfolded. It is a personal empirical timeline based on data from the Directorate-General of Health (DGS – *Direção Geral de Saúde*) regarding the number of cases and deaths in Portugal. The timeline also includes information from Portuguese social media (such as television and newspapers) concerning the main social impacts, particularly on education, as well as relevant legislation and legal context. This timeline aims to offer an understanding into the dramatic Covid-19 pandemic outbreak and its rapid propagation, from the first day of 2020 until the last day of classes in the 2019/2020 academic year.

2.1.1. Timeline of the Context

Date	Cases	Deaths	Observations
01 Jan	-	-	In China, the market believed to be the source of a new and unknown contamination is closed, as all the patients were linked to the location.
04 Jan	-	-	National authorities in China report to the World Health Organization (WHO) that there are 44 cases of patients with a pneumonia of unknown origin.
11 Jan	-	-	Chinese authorities identify the causative agent of the pneumonia cases as a new type of coronavirus, which was isolated in seven patients.
13 Jan	-	-	First confirmed case outside China is reported in Thailand.
15 Jan	 -	-	First public statement from the Portuguese authorities about the new coronavirus: The Portuguese authorities believe that the outbreak is contained, and that widespread transmission is not 'a scenario being considered at the moment.
22 Jan	-	-	Macau confirms its first case of the disease, at a time when more than 440 people are infected in the world.
23 Jan	-	-	The WHO convenes its emergency committee in Switzerland to assess whether the outbreak constitutes a Public Health Emergency of International Concern. Still, decides not to declare it.
24 Jan	-	-	France confirms the first two cases in Europe, both imported.
25 Jan	-	-	First suspected case in Portugal is reported, but tests reveal it is negative.
27 Jan	-	-	First confirmed death in Beijing. The European Centre for Disease Prevention and Control urges EU member states to adopt "rigorous and timely measures".
28 Jan	-	-	The European Civil Protection Mechanism is activated, at France's request, for the repatriation of French citizens from Wuhan, China.
29 Jan	-	-	At least 17 Portuguese citizens request to leave China, almost all from the Wuhan region.
30 Jan	-	-	The WHO declares the outbreak a Public Health Emergency of International Concern (WHO's highest level of alarm) but opposes to travel or trade restrictions.
31 Jan	-	-	The United States of America decides to ban entry for foreigners who have been in China in the last 14 days.
02 Feb	-	-	Eighteen Portuguese citizens and two Brazilians citizens evacuated from Wuhan arrive in Lisbon and start a voluntary quarantine for 14 days.
11 Feb	-	-	The WHO decides to officially name the infection caused by the new coronavirus as Covid-19.
14 Feb	-	-	Second death confirmed outside China, in Japan.
21 Feb	-	-	Italy records its first fatality, a 78 year old man.
22 Feb	-	-	Iran closes schools, universities, and educational centres in two cities.
23 Feb	-	-	Japanese authorities confirm that a Portuguese national, an employee of a ship docked at the port of Yokohama, tested positive for the Covid-19 virus.
24 Feb	-	-	The WHO warns that the world must prepare for a possible pandemic.

Table 7. Main events of the Covid-19 pandemic outbreak in Portugal

Table 7. Main events of the Covid-19 pandemic outbreak in Portugal (Cont.)

Date	Cases	Deaths	Observations
26 Feb	-	-	Several European countries confirm their first cases. The WHO reveals that the number of newly confirmed daily cases outside China has, for the first time, surpassed those reported in China.
01 Mar	-	-	The Asturian government confirms the first case of infection in this Spanish region: The Chilean writer Luis Sepúlveda, who recently visited Portugal.
02 Mar	2	-	Portugal confirms its first two cases. The Portuguese government issues a directive ordering public services to develop contingency plans for the Covid-19 outbreak (See Order No. 2836-A/2020).
03 Mar	2	-	The Portuguese government gives to public institutions, including schools, five days to develop contingency plans.
04 Mar	5	-	Italy, the most affected country in Europe, closes all schools and universities until mid-March.
07 Mar	20	-	A school in Felgueiras, the Abel Salazar Institute in Porto, the Faculty of Pharmacy from the University of Porto, and the History building at the University of Minho are closed due to their association with cases of infected individuals.
08 Mar	 30	-	DGS closes schools and suspends leisure and cultural activities in the municipalities of Lousada and Felgueiras due to the accumulation of cases.
09 Mar	30	-	The Universities of Lisbon and Coimbra suspend all in-person classes for two weeks.
10 Mar	41	-	The Portuguese government suspends flights to all regions of Italy for 14 days.
11 Mar	59	-	The WHO declares the Covid-19 outbreak a global pandemic.
12 Mar	59	-	The Portuguese Prime Minister announces the closure of all schools, starting March 16, and the closing of nightclubs, restrictions on restaurants, shopping centres and public services. The state of alert is declared (See Order No. 3298-B/2020).
13 Mar	112	-	The WHO declares that Europe is the epicentre of the pandemic.
15 Mar	245	-	Order No. 3301-B/2020 introduces exceptional and temporary measures regarding the suspension of teaching.
16 Mar	331	-	Schools close and begin Emergency Remote Teaching (ERT) as best as they can. The Ministry of Education assures that schools have conditions to carry out the 2 nd term final summative assessments.
17 Mar	448	1	Parents of students in private institutions begin to challenge the payment of tuition fees for educational services that are not being provided.
18 Mar	448	2	Parents working from home despair over the excessive amount of homework sent to their children. The state of emergency is declared by the President of the Republic (See Decree of the President of the Republic No. 14-A/2020).
21 Mar	1 280	12	The Government states that no changes to the academic calendar or curricular plans are expected, and the reopening of schools on April 9 is being considered.
27 Mar	4 268	76	Schools' 2 nd term ends and Easter break starts. Schools prepare themselves for a 3 rd term with ERT.
02 Apr	9 034	209	Renewal of the state of emergency until April 17 (See Decree of the President of the Republic No. 17-A/2020).
04 Apr	10 524	266	The Government evaluates reactivating <i>Telescola</i> [distance education through the TV] up to the 9 th grade and resuming in-person classes only for secondary education.
09 Apr	13 956	409	Government decides: <i>Telescola</i> will begin broadcasting on April 20. Most external assessments will be cancelled. Secondary education exams will be postponed. The possibility of reopening schools for secondary education students will be considered.

Table 7. Main events of the Covid-19 pandemic outbreak in Portugal (Cont.)

Date	Cases	Deaths	Observations
13 Apr	16 934	535	Decree-Law No. 14-G/2020 is approved, establishing exceptional and temporary measures in the field of education.
14 Apr	17 448	567	Restart of classes after the Easter break through distance education (ERT).
	18 841	629	Luis Sepulveda dies as a victim of Covid-19. The state of emergency is renewed until May 2 (See Decree of the President of the Republic No. 20-A/2020). UNESCO declares that "this crisis will change education forever".
17 Apr	19 022	657	Some private education institutions claim that their viability is at stake and demand governmental financial support.
20 Apr	20 863	735	#Estudo em Casa begins its broadcast at RTP Memória TV channel.
22 Apr	21 982	785	Increasing protests against restriction measures. The government considers whether to ease the measures or extend the state of emergency. Experts warn of the danger of a second wave.
25 Apr	23 392	880	The April 25 th celebrations on Avenida da Liberdade are portrayed on the Media by an old man walking alone down the avenue with a Portuguese flag.
27 Apr	24 027	928	The government is considering in-person classes for 11 th and 12 th grades starting May 18, and for preschool starting June 1.
29 Apr	24 505	973	The state of emergency will not be renewed, and a state of Calamity will be declared.
03 May	25 282	1 043	State of Calamity is declared (See Council of Ministers' Statement on April 30, 2020).
06 May	26 182	1 089	Government issues an announcement with conditions for the return to in-person classes for grades 11 and 12.
18 May	29 209	1 231	11 th and 12 th grade students return to in-person classes for subjects with exam. <i>Creches</i> [Daycare centres] reopen.
21 May	29 912	1 277	The Ministry of Education announces that the final year summative assessments will be audited to control grade inflation.
01 Jun	32 700	1 424	Pre-schools reopen.
 15 Jun	37 036	1 520	<i>Centros de Atividades de Tempos Livres</i> [After-School Activity Centres] reopen. Children over 10 years old are required to wear masks.
26 Jun	40 866	1 555	End of classes for all students in ERT, and summer break starts.
Note: Re	gular Rule	of Law	State of Alert State of Emergency State of Calamity

Sources: Antena Livre (2020); Correio da Manhã (2020); Diário de Notícias (2020); DGS (2020); Jornal de Notícias (2020)

2.1.2. Emergency Remote Teaching

On March 16 of 2020, with Portuguese schools on the final phase of 2nd term and the country in a state of alert due to the Covid-19 pandemic, the Portuguese government determined the closure of all schools across the country.

Although the closure of schools had already happened due to other large-scale epidemics (Almeida, 2014), the digital age in which we live has allowed schools to continue their work remotely. This resulted in the transfer of approximately 1.5 million students and 135 thousand teachers (DGEEC, 2021) from face-to-face to distance learning. It should be noted at the outset that this type of distance learning was not prepared or planned in advance. Nor has it been implemented by professionals with the necessary digital and pedagogical skills for distance learning (OECD, 2020b; CNE, 2021a). This solution arose as a response to a social emergency, and the new established reality was, in fact, an improvised and temporary solution. Thus, this Emergency Remote Teaching (ERT) cannot, and should not, be confused with regular distance learning, which has a solid construction such as, for example, the distance learning carried out by *Universidade Aberta* in Portugal.

As in many other countries, this transfer from face-to-face teaching to ERT has imposed profound changes in the teaching and learning processes, with manifest limitations at various levels, which several national and international studies have highlighted. In this regard, see, for example, Flores & Gago, 2020; CNE, 2021a; OECD, 2021. This shift from face-to-face teaching to ERT, a crisis answer with a wide variety of educational solutions, has also led to an evident increase in social inequalities among students (Martins, 2020). These inequalities among students were of diverse depth and nature, namely in terms of different housing conditions and different levels of parental support for remote learning (Esteves et al., 2021), different levels of access to technological equipment and to the internet (OECD, 2020b), and different levels of students' self-efficacy, self-regulation, and motivation (Ikeda & Echazarra, 2021).

The inequalities in access to ERT and the disparities in the quality of ERT had an impact on students' learning, on students' performance and on students' assessments. In an equitable assessment system, assessments are fair, reliable, and valid. However, in a context where the fairness, reliability, and validity of assessments are widely questioned, the use of assessment for accountability purposes became increasingly controversial as, ultimately, these inequalities would have an impact on decisions based on these assessments (NAE, 2021).

Consequently, also in Portugal as around the world, concerns immediately arose about the pandemic's impact on learning and summative assessments, as well as on conditions for national exams and access to higher education. The study conducted by the *Observatório de Políticas de Educação e Formação* showed that, in April 2020, among secondary education students in Portugal, there was a high level of concern about final year grading and higher education access. More than half of the surveyed students were very concerned, while only 12% were not worried about their summative assessments (Benavente et al., 2020).

To mitigate the pandemic's effects on the Portuguese education system and to address these and other concerns regarding summative assessment processes, Decree-Law No. 14-G/2020 was issued in April 2020, introducing amendments to Decree-Law No. 55/2018.

2.2. The Decree-Law No. 14-G/2020: An Emergency Assessment Policy

As referred to in Chapter 1, the curriculum for basic and secondary education, along with the guiding principles for its design, implementation, and assessment of learning, is established through Decree-Law No. 55/2018. However, in the 2019/2020 academic year,

(...) as there [were] situations requiring explicit regulation within this exceptional scope due to the evolving Covid-19 pandemic, the Government decided to approve a set of measures in the field of education aimed at establishing an exceptional and temporary regime regarding assessment of learning, the school and examination calendars for basic and secondary education, enrolments, registration for national final exams, and teaching and non-teaching staff, in order to ensure the continuity of the 2019/2020 school year in a fair, equitable, and as normalized manner as possible (Decree-Law No. 14-G/2020).

Thus, the enactment of Decree-Law No. 14-G/2020 in the field of education is not intended to correct or repeal Decree-Law No. 55/2018, but rather to address the need to regulate teaching and assessment processes in an emergency context. In this sense, the changes introduced on teaching and assessment processes are, by definition, exceptional and temporary, only valid for the 2019/2020 academic year. Additionally, Decree-Law No. 14-G/2020 only adjusts the school calendar and school administration processes to accommodate the changes that were introduced in teaching and assessment processes for the conclusion of the 2019/2020 academic year.

The following sections of this thesis chapter focuses on student assessment policy in an emergency context, and the implications of Decree-Law No. 14-G/2020 and ERT on both internal and external summative assessment processes for the conclusion of the 2019/2020 academic year.

2.2.1. Student Internal Assessment Policy in ERT

During ERT, when changes to internal assessment processes were introduced through Decree-Law No. 14-G/2020, no reference is made to formative internal assessment processes. Throughout this Decree-Law, whenever there are references to internal assessment processes, its wording reflects concerns primarily centred on summative assessment processes, students' final grading, and certification of learning. It is true that internal summative assessments can also facilitate students' learning when it involves active participation in processes of analysis and reflection on the learning achieved (Fernandes, 2011; NRC, 2001a). However, in practice, summative assessments are more closely associated with grading, certification, and selection (Fernandes, 2011); a conception and use that is very evident in the wording of Decree-Law No. 14-G/2020 for the conclusion of the 2019/2020 academic year. Recommendations and guidelines for formative classroom assessment and the provision of feedback in ERT were specified through guidelines developed by the Ministry of Education, incorporating guiding principles for pedagogical assessment in distance learning¹.

Note that with the enactment of Decree-Law No. 14-G/2020 – written with a clear focus on the use of assessments for grading and certification of general basic education, secondary education, specialized artistic courses, and other educational offerings – only internal assessment should be considered for summative purposes. In addition, it is worth noting that the grades assigned in each subject area should have as a reference all learning activities completed by the end of the academic year. This should include the work done during the 3rd term in ERT, without prejudice to the overall assessment of students' learning.

It is also important to highlight that, according to the guidelines of the Portuguese Ministry of Education, schools continued to have autonomy to define the distance learning methodologies they considered most appropriate. This involved pondering all available resources, assessment criteria, and students' living contexts, aiming to establish an equitable access to learning and assessment.

To all students it was also required to maintain attendance in synchronous sessions, when available, and when students had the technological tools to access them. Furthermore, students were required to complete the activities proposed in asynchronous sessions and submit completed tasks so that teachers could gather evidence for final summative assessments.

¹ See, for example, <u>https://www.dge.mec.pt/sites/default/files/roteiro_avaliacao_ensino_a_distancia.pdf</u>

2.2.2. Student External Assessment Policy in ERT

With the approval of Decree-Law No. 14-G/2020, all PA, PFC and EN were cancelled for final grading and learning certification. Students in secondary education only took final EN in subject areas they chose for higher education access. Students had also the option to retake exams for grade improvement, but results would only be considered for higher education access scoring.

In the 2019/2020 academic year, the following tests are cancelled: a) PA for the 2nd, 5th, and 8th grades; b) PFC, at the end of the 9th grade; c) School-level tests conducted as final exams for basic education; d) EN, when taken by internal students, for the purpose of subject areas approval and completion of secondary education (Article 6 of Decree-Law No. 14-G/2020).

Furthermore, in June 2020, IAVE published an addendum to EN information regarding external assessment in secondary education. In this school year of 2019/2020, EN would include sets of mandatory and optional items, allowing all students to engage with items covered in classes. Students could respond to all items as usual, as the computer system would automatically select only the responses to items where students demonstrated better performance, in addition to responses to items that mandatorily contribute to the exam final marking (IAVE, 2020).

Given the emergence of this first lockdown and the critical issue regarding the real learning opportunities students had, it is understandable that the Portuguese government's prime concern, in cancelling most tests and exams, was to mitigate the pandemic's effects on student accountability. Efforts were simultaneously made to minimize the impact of final exams on decisions affecting students. However, it is important to consider that external assessments not only serve to measure what students know and are capable of doing. They also serve to collect information to understand the performance of the education system, evaluate and improve learning and teaching, and contribute to enhancing the quality of education (Fernandes, 2019), information of utmost importance in a disruptive context like that caused by the Covid-19 pandemic.

On the other hand, assessments are only useful if those who can benefit from their information can access, interpret, and use that information with guarantees of fairness, validity, and reliability (Stufflebeam, 2014). Thus, faced with this dichotomy of whether to use external assessments or not, the pandemic brings to the forefront the importance of understanding and documenting learning processes and contexts and the need to consider them in the design and interpretation of external assessments (NAE, 2021).

2.3. Early impacts of ERT on Student Assessment

2.3.1. ERT Impacts on Internal Assessment

Regarding early impacts on internal assessment practices, the study conducted by CNE (CNE, 2021a), titled *Educação em Tempo de Pandemia: problemas, respostas e desafios das escolas* [Education in Times of Pandemic: problems, responses, and challenges of schools], reveals that 82% of the 4 338 teachers with coordination functions found assessing learning to be difficult or very difficult during ERT. To assess learning, the surveyed teachers indicated that they had to reformulate methods, instruments, and criteria, demonstrating a commitment to finding ways to value students' work and evidence of learning in an atypical context. To this end, 50% of teachers indicated that they had changed the way they collected information to assess learning. Moreover, 26% of the surveyed teachers indicated that they had made changes to the assessment criteria, and 31% introduced changes to the purposes and uses of assessments, prioritizing formative purposes.

(...) the practices, if not the conceptions, of teachers regarding assessment, were confronted with changing circumstances. Distance raised questions regarding methods, instruments, criteria, objects, and trust (CNE, 2021a: 147).

Regarding the modification of assessment criteria, it is important to recall a point made by the National Council of Education:

In this normative [Decree-Law No. 14-G/2020], the objects to be assessed are not restricted. Therefore, "valuing more the attitudinal domain (participation, interest, commitment, attendance, timely completion of tasks, ...), valuing less the domain of knowledge," as testified by a teacher, does not come from a nationally established guideline, but rather from a decision made by the school. It reflects, therefore, the chosen learnings or the specific circumstances of that school (CNE, 2021a: 151).

Concerning assessment instruments and collected information used by teachers as the assessed object during this period, another study conducted by the *Centro de Economia da Educação* at Nova SBE revealed that the majority of their surveyed teachers (84%) relied on homework assignments for

their assessment processes. Over two-thirds (68%) of teachers also used attendance in synchronous sessions as an assessed element, and participation in videoconference classes was used by 65% of teachers. Only approximately one-third of teachers, about 33%, performed assessments through tests (Reis et al., 2020). Additionally, Rodrigues et al. (2022) shows that teachers found oral discussions, dialogue simulations, and online presentations the most suitable assessment methods, as they were easier to implement compared to traditional tests, peer reviews, text processors, and quizzes.

Regarding to these changes to internal assessment practices during the initial phase of ERT in Portugal, it is essential to acknowledge that despite the principles outlined in public policies, ministerial recommendations, teacher professional development programs, head teachers' guidelines, or instructions from pedagogical councils, the pandemic and distance have profoundly altered the context and the pedagogical dynamics between teachers and students. The type of interaction that a teacher establishes with the students is directly related to the data collection procedures, the purposes of that assessment, the stakeholders' interests in the results of that assessment, and the mode of results dissemination (Fernandes, 2010, 2013; Stufflebeam & Coryn, 2014). Consequently, given the altered pedagogical relationship in the teaching and learning process, adjustments in assessment practices were inevitable.

Secondly, in line with the previous point, the use of new and different assessment instruments during ERT, as emphasized by some studies (e.g., CNE, 2021a; Reis et al., 2020; Rodrigues et al., 2022), reflects the altered contexts and relationships established between teachers and students. These changes on assessment instruments also demonstrate a shift in the use of data collected by teachers, who prioritized formative assessment and feedback over using this information for grading and certification purposes.

Thirdly, in addition to the direct impact of the implementation of ERT on the teacher-student pedagogical relationship and internal assessment practices, it is essential to consider the impact of public policies produced during this period. While enacted by governments, public policies involve various actors and decision-making levels (Souza, 2006; Santiago et al., 2012). Moreover, the guidance provided by public policies is often ambiguous or unclear, leading to varied interpretations and implementations in different contexts, which do not always translate into direct, coherent, and obvious practices (Ball, 2008). The variety of educational solutions adopted by schools, as well as the diverse assessment practices implemented by teachers, also reflect the various interpretations and implementations of policies when *"put into practice"*, and the chosen learnings or the specific circumstances of a school (CNE, 2021a).

Lastly, when still considering internal assessment practices, it is crucial not to overlook other influencing factors on teachers' approaches. It is extremely relevant to consider a range of other factors, from teacher training programs like the MAIA project implemented during the pandemic

(Fernandes et al., 2021), to head teachers' guidelines, advices from school pedagogical councils, and even strictly personal factors that may have influenced teachers' practices. Therefore, considering this disruption in teachers' assessment practices, a first research question (RQ) arises:

RQ1: Has changing assessment methodologies facilitated teachers' adaptation to Emergency Remote Teaching?

2.3.2. ERT Impacts on External Assessment

With the enactment of Decree-Law No. 14-G/2020 and the subsequent cancellation of PA, PFC, and EN for subject approval and secondary education conclusion, students only participated in final EN for subjects chosen as admission exams for higher education access. Furthermore, students in secondary education were allowed to retake these exams for grade improvement, with the results being considered only into higher education admission exam scores.

As EN were no longer mandatory for secondary education conclusion, despite the total number of enrolled students in secondary education remaining nearly constant compared to the previous two academic years — with approximately 350 000 students enrolled across all secondary education levels (10th, 11th, and 12th grades) — there was a significant decrease in the number of exams taken, as shown in Figure 5. The most notable decline was observed in Portuguese exams, which were mandatory for all scientific-humanistic courses until 2019 but became optional in 2020.



Figure 5. Number of students enrolled in secondary education and number of exams taken per academic year (2018-2020)

Sources: DGE, 2021; DGEEC, 2021

According to data from DGEEC (2023) illustrated in Figure 6, the conclusion rate of secondary education (12th grade) in 2020 was 84.6%, representing an increase of 7.2 percentage points compared to 2019. This increase is the highest positive variation in conclusion rates observed in the time period from 2010 to 2022, and it also corresponds to the highest conclusion rate of secondary education ever recorded in Portugal until then.





Note that the removal of EN reduced some of the pressure these exams placed on secondary education conclusion, and students experienced greater separation between completing secondary education and accessing higher education.

One of the most detrimental effects of external assessments is the so-called washback effect, in which teachers tend to align their teaching and assessment practices with the content and methods they believe will be covered and tested in exams. This tendency increases when exam results carry greater consequences (Fernandes, 2019; Madaus & Russell, 2010). Therefore, with fewer external exams and fewer decisions in secondary education based on exams' results, there was also less pressure exerted by exams on secondary education. This pressure relief had been recommended by the CNE even before Decree-Law No. 14-G/2020 was approved, with the goal of ensuring that secondary education better served its intended role as a terminal stage (Lourtie, 2020).

Source: DGEEC, 2023 through EDUSTAT

The weight of national exams, along with the pressure from candidates for higher education admission to prepare for these exams, has negative effects on the pursuit of the objectives of secondary education. This contradicts its intended role as a terminal education, which is aimed at preparing young people for life as active and conscientious citizens in a democratic society (Lourtie, 2020: 2).

With a larger number of students completing secondary education, there was also a larger number of students eligible to access higher education. This resulted in a significant increase in applicants for higher education, as indicated by the study conducted by EDULOG - Belmiro de Azevedo Foundation, titled *"Estudantes nacionais e internacionais no acesso ao ensino superior: quem são, que escolhas fazem e como acedem ao mercado de trabalho"* [National and International Students in Higher Education Access: Who they are, what choices they make, and how they access the job market]:

On September 3rd, 2020, during the National Higher Education Admission Process, Decree-Law No. 62-A/2020 was approved. This decree enabled institutions to transfer vacancies not filled in the special admissions regime to the general access regime. The decision aimed to increase the number of vacancies in public higher education in response to the significant increase in candidates during the first phase of the National Higher Education Admission Process (Sá et al., 2021: 38).

It is important to note that this increase does not solely reflect more students completing secondary education. In addition to national exams being conducted solely for higher education admission purposes, the exams administered in 2020 had a different test structure compared to previous academic years. Data from the Directorate-General of Education, presented in Table 8, demonstrate that in 2020, the variation in average scores increased by almost 19% compared to the previous academic year.

Table 8. Average scores in EN (0-200 points) and year-on-year variation rate (%) from 2015-2020

	2015	2016	2017	2018	2019	2020
Average scores in EN	108	105	109	110	107	127
Year-on-year variation		-2.3%	+3.6%	+1.0%	-2.5%	+18.9%

Source: DGE, 2021; Own calculations

This represents a much larger fluctuation than in previous academic years, and this positive variation has created better conditions for the use of national exams as a higher education admission test. Consequently, not only did more students complete secondary education, but more students also succeeded in the higher education entrance exams. The combination of these factors led to the highest number ever of students applying (62 561 students) and accessing (50 964 students) higher education in Portugal during the first phase of the national access process in 2020 (DGES, 2021).

It is important to note that the strategies adopted to handle examinations during ERT have varied from country to country, as shown in Table 9.

Changes	Countries	Ν
Introduced additional health and safety measures (e.g., extra space between desks)	Austria, Belgium (Flemish), Belgium (French), Chile, Colombia, Czech Republic, Estonia, Finland, France, Germany, Hungary, Israel, Italy, Lithuania, Latvia, Poland, Portugal, Russian Federation, Slovenia, Spain, Turkey	21
Adjusted the content of the examinations (e.g., subjects covered or number of questions)	Austria, Chile, Spain, Israel, Italy, Latvia, Poland, Portugal, Russian Federation, Turkey	10
Adjusted the mode of administration (e.g., computer-based or online-based)	Belgium (Flemish), Colombia, Italy, Latvia, Lithuania	5
Postponed/rescheduled the examinations	Austria, Chile, Colombia, Czech Republic, Germany, Spain, Estonia, Finland, Israel, Korea, Latvia, Lithuania, New Zealand, Poland, Portugal, Slovenia, Turkey	17
Cancelled the examinations and used an alternative approach for high-stakes decision making (e.g., calculated grades)	Belgium (French), Denmark, Estonia, France, Hungary, Israel, Netherlands, Norway, Slovak Republic	9
Introduced alternative assessment/validation of learning (e.g. appraisal of student learning portfolio)	Costa Rica, France, Israel, Latvia, Netherlands, New Zealand, Poland, Russian Federation	8

Table 9. Changes to 2019/2020 national examinations due to the Pandemic within OECD countries

Note: 34 countries completed the questionnaire. Of these, 28 provided valid answers to at least one of the questions. Source: OECD/UNESCO-UIS/UNICEF/World Bank Special Survey on Covid: March 2021 – Through OECD (2021:29)

Still, the solutions found by the Portuguese government for the conclusion of secondary education and access to higher education through the general access regime seem to have been widely accepted by school and academic communities. Perhaps, for this reason, in 2021, 2022 and 2023, unlike some countries, the solutions implemented by the Portuguese government were very similar to those of 2020, with the approval of Decree-Law No. 22-D/2021 in 2021; the approval of Decree-Law No. 27-B/2022 in 2022; and the approval of Decree-Law No. 22/2023 in 2023. In the United Kingdom, for example, in 2020, with the cancellation of *A-levels*, grades were calculated based on a performance prediction algorithm, based on teachers' past assessments and normalized according to each school's historical performance. However, these grades generated much controversy across the country. Therefore, in 2021, the British government was forced to adopt different solutions (Nisbet & Shaw, 2022; Sá et al., 2021: 32-34).

Thus, these changes, in Portugal and worldwide, may constitute interesting starting points for an evaluation of higher education access systems, both in their effectiveness and suitability, and also in their contribution to a more inclusive higher education system (Sá et al., 2021: 29). Since little is yet known about how this temporary suspension affected schools' functioning in Portugal, this leads us to the second research question:

RQ2: How was the cancellation of external assessments perceived in schools?

2.3.3. ERT Impacts on Grade Repetition

Although hypothetically temporary, teachers' assessment practices during ERT, as examined in 2.3.1, contradict beliefs regarding the purposes and uses of assessment, particularly concerning grade repetition. Grade repetition is a deeply ingrained practice in the Portuguese teaching profession, rooted in the use of assessment for grading and selection purposes, a practice that supersedes national laws and scientific recommendations (Cipriano & Martins, 2021).

Despite evident learning losses experienced by students during ERT (see, e.g., Engzell et al., 2020; IAVE, 2021; Maldonado & De Witte, 2020; Tomasik et al., 2020), in 2020, grade repetition rates in Portugal have reached historically low levels during the pandemic first year when comparing to its precedent years. Data from the *Direção-Geral de Estatísticas da Educação e Ciência* [Directorate-General of Statistics for Education and Science] (DGEEC) demonstrate that in the year of 2020, grade repetition and dropout rates deepened a trend that had already been observed during the period under analysis (Figure 7). Additionally, in Table 10, when calculating the year-on-year variation rate, it is observed that from 2019 to 2020, certain grade levels, for the first time, experienced a variation greater than 3 percentage points.



Figure 7. Grade repetition and dropout rates, by academic year and grade level (%)

Source: DGEEC, 2023 through PORDATA

Table 10. Year-on-year variation of grade repetition and dropout rates, by academic year and grade level (pp)

	2016	2017	2018	2019	2020	2021	2022
Grade 1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Grade 2	-0.4	-1.5	-0.8	-1.7	-1.7	+1.0	-0.5
Grade 3	-1.1	-0.8	0.0	-0.7	-0.5	+1.0	-0.4
Grade 4	0.0	-0.4	-0.1	-0.3	-0.3	+0.7	0.0
Grade 5	-1.7	-0.7	-0.5	-1.6	-1.5	+0.5	0.0
Grade 6	-2.0	-1.1	-0.6	-1.3	-1.4	+1.2	-0.2
Grade 7	-2.8	-1.2	-1.6	-2.8	-2.8	+1.5	0.0
Grade 8	-2.3	-1.3	0.1	-2.1	-2.0	+1.5	-0.1
Grade 9	-1.7	-2.0	-0.5	-1.0	-3.3	+0.6	+0.9
Grade 10	+0.4	-0.3	-1.5	-0.4	-3.4	+1.1	+0.9
Grade 11	-1.5	0.0	-0.1	-0.3	-3.4	+0.4	+0.6
Grade 12	-1.7	-2.0	-1.7	-1.9	-7.2	-2.5	-0.7

Source: DGEEC, 2024

The grade levels that saw the highest variation in grade repetition and dropout rates were the 9th grade, within basic education, and the 12th grade, within secondary education; with negative variations of 3.3 and 7.2 percentage points, respectively. It is worth noting that these grade levels, the 9th and 12th grades, unlike in previous academic years, were no longer dependent on external national exams for completion and certification of general basic education and secondary education. Their completion was now solely dependent on internal assessment.

Since grade repetition practices were so affected in the 2019/2020 academic year and they relied only on teachers' decisions, a third research question (RQ) arises:

RQ3: Did the Covid-19 pandemic affected teachers' grade repetition beliefs?

2.3.4. ERT Impacts on Digital Transition

Within ERT, schools and teachers were responsible for defining and implementing a distance learning plan to ensure that all students had access to learning.

With the suspension of in-person teaching activities in schools, learning is to be carried out through remote teaching, using the methodologies that each school deems most appropriate (...). It is the responsibility of schools, with the support of the Ministry of Education central services and in coordination with entities acting as partners, to implement the distance learning plan, ensuring that teachers monitor students of each class, aiming to provide an equitable access to learning for all (Decree-Law No. 14-G/2020).

In this context, a wide variety of solutions emerged from school to school, and most schools decided to implement a distant teaching plan using synchronous sessions through the online platforms Google Meet and Microsoft Teams, while fewer schools also used Skype and Zoom. Additionally, teachers sought to maximize the potential of the Moodle platforms that were already in place before the Covid-19 pandemic (Flores & Gago, 2020), and Schoolbook publishers also provided free access to teaching platforms with strategies and suggestions (Costa & Baptista, 2023), such as *Escola Virtual* from *Porto Editora*.

However, teachers faced numerous challenges, including reliance on personal devices for teaching, inadequate equipment for students, insufficient training in online teaching, limited time, and a lack of parental support (Flores & Gago, 2020). The lack of resources was indeed a major concern for teachers, which disproportionately impacted disadvantaged students and contributed to social inequalities and exclusion. This shortage of students and teachers digital resources and literacy also hindered teachers' work, highlighting the need for investment in students' access to technology and stronger institutional support for teachers (Flores et al., 2023).

Given this scenario, on April 21, 2020, a Digital Teacher Training Plan was launched by the Directorate-General for Education (DGE) in collaboration with Teachers Training Centres. The plan aimed to enhance teachers' digital skills, included three training levels considering teachers proficiency. Teachers participated in training and additional initiatives according to their school's strategy (Costa & Baptista, 2023). Additionally, also in April 21 of that year, the Council of Ministers Resolution No. 30/2020 approved the Digital Transition Action Plan for public administration, businesses, and citizens in general. Regarding education, this plan includes:

- The provision of individual equipment to each student, tailored to the needs of each educational level, for use in a learning context;
- The provision of free mobile connectivity for students and teachers;
- Access to quality digital educational resources;
- Access to collaboration tools in digital environments that promote innovation in the teachinglearning process;
- The establishment of processes leading to the electronic marking of external assessment tests in a digital environment.

The plan also places a strong emphasis on teacher training through a digital skills development program, ensuring that teachers acquire the necessary competencies for teaching in this new digital context. It aims to actively contribute to the technological modernization of schools, bringing students closer to the productivity and collaboration tools they will encounter in a professional work environment (Council of Ministers Resolution No. 30/2020: 14-15).

Considering the mandatory use of technology during the pandemic, the fact that the digital tools adopted during the pandemic brought innovation to classrooms practices and most of them remained in a post-pandemic context (Costa & Baptista, 2023), and considering the planned governmental investments for the digital transition in education, the final research question arises:

2.4. Chapter Final Considerations

The institutional data provided by DGE, DGEEC, and the aforementioned studies indicate a substantial shift in assessment policies and practices during the initial phase of the pandemic in Portugal, along with corresponding decision-making processes. However, attempting to attribute these changes to a single cause/effect relationship or categorize them as either temporary or permanent oversimplifies the situation, as education systems are dynamic social constructs, constantly evolving and influenced by numerous factors.

Nevertheless, the Covid-19 pandemic has undeniably prompted significant changes in the inputs received by the education system, resulting in corresponding adjustments in its outputs. As a result, the social disruption caused by the pandemic has sparked a constructive discussion about the education system's capacity for adaptation and resilience in identifying challenges and devising solutions amidst adversity. Moreover, it has prompted reflections on the future design and performance of education systems, particularly in a post-pandemic context (OECD, 2020a).

The pandemic and the confinement period have also ignited a global debate on assessment practices (Martins, 2020). At the heart of this discourse lies a fundamental and recurring question: What is the purpose of assessments and to what extent do they improve teaching, learning, and mitigate educational disparities (NAE, 2021). This period also shed light on the role of external assessments in education systems. The forced introduction of new conditions for the conclusion of secondary education as for the access to higher education in Portugal, deserves serious reflection on the potential and limitations of EN, and on which higher education access system we want in a post-pandemic era.

Finally, it is crucial to recognize that summative assessments of students' learning have historically been associated with stratification mechanisms, significantly impacting educational paths and opportunities (Martins, 2017). Despite the undeniable learning disruptions caused by the pandemic and ERT, changes in summative assessments and on the decisions based on these assessments appear to have mitigated their consequences on students' educational paths. In 2020, Portugal recorded its lowest ever rates of grade repetition and school dropout, alongside the highest number of students completing compulsory education and accessing higher education. However, the data from these forced changes deserve further study on their real impact on mitigating disparities in students' paths and opportunities. Thus, the pandemic and the ERT period have brought opportunities, with real cases and real data, that question the purposes and use of internal assessment, as well as changes in the purposes and use of external assessment; opportunities that, due to the high social and human cost of this pandemic, cannot be wasted, and are explored in the following chapters of this thesis.

CHAPTER 3

Research Design

Chapter Summary:

In this Chapter 3, Section 3.1 details the taxonomy development and approaches assignment for research in this thesis. Since it was decided to address each research questions and objectives independently, the following sections of this chapter (Sections 3.2, 3.3, 3.4, and 3.5) specifies the research approaches for each sub-study, considering each sub-study objective, data sources, sampling processes, inclusion/exclusion criteria and participants characteristics. Additionally, the development of instruments and procedures for primary data collection are also explained, along with data analysis techniques clarification. Section 3.6 is dedicated to research ethical issues.

3.1. Taxonomy Development and Approach Assignment

Concerning the impacts of the Covid-19 pandemic on student assessment policy and practice, four research questions emerged in the previous Chapter 2:

- RQ1: Has changing assessment methodologies facilitated teachers' adaptation to ERT?
- RQ2: How was the cancellation of external assessments perceived in schools?
- RQ3: Did the Covid-19 pandemic affected teachers' grade repetition beliefs?
- RQ4: Are schools ready for DAVE's implementation?

To address these research questions and to clarify the studied object, the taxonomy² categories that serve as a reference for the study are outlined in the following points of this section:

- 1. Internal Assessment Practices in ERT
- 2. Cancellation of External Assessments
- 3. Grade Repetition Beliefs
- 4. DAVE's Implementation

² The term taxonomy comes from the Greek *taxis* and *nomos*, and it refers to the *rules or conventions of order or arrangement*, i.e., the structured names and definitions used to organize information and knowledge (Lambe, 2007:4).

Note that, to answer these research questions of different natures, it was decided to address each of them independently. This approach allows for an in-depth study in which both quantitative and/or qualitative research methods are used to address the various research questions (Bryman, 2012: 640). Still, all objectives were defined with consideration of i) the Covid-19 context, ii) student assessment policies, and iii) student assessment practices.

3.1.1. Internal Assessment Practices in ERT

Several studies have shown that meaningful classroom assessment is strongly linked to the curriculum, instruction and learning processes (e.g., Black & Wiliam, 1998a, 2018; Fernandes, 2011; Lingard, 2007). In March 2020, the implementation of ERT due to the Covid-19 pandemic brought strong limitations to instruction and learning processes, with strong implications to assessment of learning (E.g., Cooper et al., 2022; NAE, 2021; Panadero et al., 2022; Rodrigues et al., 2022; Seabra et al, 2021). In Portugal, these ERT processes were regulated by Decree-Law No. 14-G/2020. Yet, in a very disruptive social context as the Covid-19 pandemic was for schools, teachers, and students, it becomes crucial to understand how have teachers deal with assessment of learning, what were the impacts of this new legal framework in their practices, and how it relates with their adaptation and perceived workloads.

Note that for special education teachers, literature recommends special education teachers to adjust learning targets, instruction, and assessment methods according to each student's needs (e.g., Alves et al, 2020; Brookhart, 2013; Lingard, 2007; Xu, 2013). Nevertheless, individualized learning targets are still related to the general education curriculum, instruction, and assessment processes (Xu, 2013). In Portugal, aiming to consolidate an effectively inclusive education in Portuguese schools, these recommendations have also been considered in most recent educational policies (see Decree-Law No. 54/2018). Hence, when it comes to special education teachers practices during this period, some studies have already showed that special education services continued to be delivered. Still, special education professionals reported lower levels self-efficacy in their roles (Womack & Monteiro, 2022) and special education teachers described a tension around accountability and grading (Schuck & Lambert, 2020).

Therefore, objective one is as follows:

Objective 1: To investigate whether the enactment of Decree-Law No. 14-G/2020 eased the challenges faced by teachers in their instructional and assessment practices, with a particular focus on teachers of vulnerable groups, such as students with Special Education Needs and Disabilities (SEND). To this end, the use of moderation and mediation models – statistical models that allow for understanding relationships between variables (Hayes, 2022) – will make it possible to analyse the relationships between different concepts. Applied to a secondary database collected immediately after the first period of ERT, the first part of sub-study 1, which focuses on teachers from all curricular areas, aims to uncover how challenging it was for Portuguese teachers to adapt to ERT and its relation to perceived workloads during this period. Moreover, it will be studied as to whether instruction and assessment mediated the relationship between teachers' own adaptation and perceived workloads. Also, it will be studied if changing assessment methodologies, as proposed by Decree-Law No. 14-G/2020, moderated the relation between teachers' own adaptation to ERT and instruction and assessment.

Regarding special education teachers, when it comes to effective assessment practices, it is known that, despite scientific recommendations to assess special education students over achievement, special education teachers tend to assess their students over improvement (Brookhart, 2013; McLeskey & Waldron, 2002). As special education teachers tend to assess over improvement and they provide individualized accommodations on assessment of learning, it will be analysed if there are differences between the whole population of teachers and special education teachers about changing assessment methodologies during this period. Thus, on the second part of this sub-study 1, focused on special education teachers only, the aim is to identify if difficulties felt by special education teachers have the same relations as when all curricular areas were considered.

3.1.2. Cancellation of External Assessments

Worldwide, in 2020, in a forced and unexpected way, the new context of emergency has also involved a shift from large-scale external assessments to teacher-built assessments, placing more responsibility on schools and teachers to leverage classroom assessment as the source for information about student growth and achievement (Cooper et al., 2022). In Portugal, considering the massive disruption lived in schools during this period (e.g., CNE, 2021a; Flores & Gago, 2020), the Portuguese government has also decided to approve extraordinary legislation for the conclusion of the 2019/20 school year (see Decree-Law No. 14-G/2020). This new legal framework has cancelled most of the pre-existing external assessment of learning moments, and very similar decisions were then pursued in the following years, with the approval of Decree-Law No. 22-D/2021 in 2021; the approval of Decree-Law No. 27-B/2022 in 2022, and the approval of Decree-Law No. 22/2023 in 2023.

In 2023, in a post-pandemic context, a full reintroduction of external assessment of learning was again discussed; and the public debate about the role of external assessment in the Portuguese education system was a hot topic of discussion in schools and academic communities. A careful

reflection regarding external assessment models beyond the Covid-19 pandemic era should consider what schools are today and what we want for/from schools in the future. Furthermore, such reflection involves an important discussion around the purposes of external assessment of learning, considering various positions of values, perspectives, and ideologies regarding learning, schools, and, ultimately, the society in which we intend to live (Fernandes, 2014b).

It is important to note that student assessment, in the Portuguese school system, operates at four key levels: education system, school, teacher, and student. Schools benefit from some autonomy in the organisation of the various components of student assessment (Santiago et al., 2012). Since 2008, with the approval of Decree-Law No. 75/2008, a one-person leadership model (head teacher) was reintroduced in the Portuguese public schools' management, in place of collegiate management. Consequently, head teachers became a key link to implement and adjust governmental student assessment policies to internal contexts and practices.

Note as well that, worldwide, it is becoming increasingly common for schools to incorporate student test scores in large-scale assessments for accountability purposes (Smith & Holloway, 2020), and in Portugal there is already a centrality of students' academic results in the daily management and accountability of Portuguese schools (Torres et al., 2019). Therefore, head teachers are an important actor to hold this reflection about the impact of suspending external assessment of learning during the Covid-19 pandemic, and the model to come in a post-pandemic context.

Thus, the second objective is:

Objective 2: To explore the implications of Decree-Law No. 14-G/2020 on the perception of school head teachers regarding the relationship between internal and external assessment of learning and its impact on school management.

To this end, in sub-study 2, it was decided to collect head teachers' points of view regarding the impacts of external assessments of learning on Portuguese schools, the impacts of its temporary suspension during the Covid-19 pandemic, and which model of external assessment of learning should be implemented for the future. It should be noted that washback and impact studies, such as this sub-study, require researchers not only to understand, but also to consider local educational context, as well as the larger social, political, and economic factors governing teaching and learning in relation to a test/examination or a testing system (Tsagari & Cheng, 2017). Consequently, to allow the emergence of these contextual factors during the research, it was decided to implement a qualitative approach

with a large set of semi-structured interviews to collect the views and opinions from the participants at this regard (Creswell & Creswell, 2023).

3.1.3. Grade Repetition Beliefs

In education, grade repetition or grade retention refers to the practice of requiring a student to remain in the same grade level for an additional school year and the student is not promoted to the next grade level along with their peers of the same age (Jimerson, 2001a). In countries where grade repetition rates are high, the main reason behind this decision is often the belief that grade repetition will provide low-achievement students a *"second chance"* to improve their learning, knowledge, and skills. However, many other factors can also influence this decision, such as students' behaviour, students' background characteristics, and cultural traditions (European Commission et al., 2020; OECD, 2023b).

Worldwide, with schools' closure during the Covid-19 pandemic and the establishment of ERT, inequities in access to education affected students' opportunities to learn (OECD, 2021). This disruptive scenario brought substantial learning losses for students (OECD, 2023b), creating great concerns regarding the validity, reliability and fairness of assessments and grading (NAE, 2021; Nisbet & Shaw, 2022). Therefore, during this period, many countries have decided to introduce exceptional regulations to internal and external assessments, grading, and grade transition conditions (E.g., CNE, 2021a; Panadero et al., 2022; Sandvik et al., 2023), placing more responsibility on teachers and schools to deal with assessments according to students' learning contexts (Cooper et al., 2022; OECD, 2021).

Despite learning losses caused by the Covid-19 pandemic, in some countries where grade repetition is a cultural practice and grade repetition rates are usually high, it was observed that grade repetition rates have considerably dropped during the 2019/2020 school year (e.g., Cipriano & Martins, 2023; Wills & van der Berg, 2024). In fact, according to the *Education at a glance 2023* report, concerning upper secondary completion rates, 20 of the analysed countries increased their completion rates during the Covid-19 pandemic (OECD, 2023a), accelerating the trend towards the reduction of grade repetition use as an educational intervention (OECD, 2023b).

Considering these facts, it is sought to understand if the Covid-19 pandemic affected Portuguese teachers' grade repetition beliefs, and whether the decline in grade repetition rates, particularly expressive in the 2019/2020 school year, are complemented by a change in teachers' grade repetition beliefs. Furthermore, it is sought to study the interaction effect of professional development on assessment literacy (the MAIA project) with teachers' grade repetition beliefs and practice. Consequently, the third objective is:

Objective 3: To identify the impacts of the Covid-19 pandemic on teachers' beliefs towards grade repetition, examining the interaction effect with the MAIA project.

To this end, a longitudinal sub-study is designed to have observations of the same phenomenon in two different time periods. More specifically, this is a trend study, which is a type of longitudinal study that examines changes within a population over time (Babbie 2013: 106). Thus, data from two independent samples, collected through two survey questionnaires in two different time periods are compared. The first sample, a secondary data base collected before the Covid-19 pandemic. The second sample, primary data collected after the Covid-19 pandemic, where a part of these teachers participated in an assessment literacy' professional development programme during the pandemic years (the MAIA project).

3.1.4. DAVE's Implementation

In recent decades, the potential benefits of introducing technologies into large-scale tests have been much discussed (see, e.g., Bennett, 2001; Kerrey & Isakson, 2000; Oldfield et al., 2012). Yet, the path to effective technology use on large-scale testing has fallen short of expectations, especially when these tests have medium or high stakes for students (Richardson & Clesham, 2021).

During the first ERT, despite many limitations, Portuguese teachers were compelled to use technology to teach remotely (Flores & Gago, 2020; Flores et al., 2023), and digital tools adopted during the pandemic brought innovation to classrooms practices, with most of them remaining in a post-pandemic context (Costa & Baptista, 2023). After a temporary cancelation of external assessment of learning due to the Covid-19 pandemic, along with other investments on digital transition in Education, the Portuguese government is taking a step forward towards the digital transition of external assessments with the DAVE project, where the reintroduction of external assessment of learning in the Portuguese education system foresees all national assessment tests (PA) and all national examinations (PFC and EN) in digital format (IAVE, 2022).

As proposed by Johnson & Shaw (2018) to study positive and negative impacts on the development and implementation of a computer-based testing initiative, with sub-study 4 it is sought to understand the perspectives of different stakeholders regarding the implementation of a computer-based large-scale testing system in Portugal. Consequently, this sub-study aims to contribute to the deepening of understanding about the development and implementation of a computer-based large-scale testing system, using Portugal as a case study. Such a study will provide a comprehensive overview of the use of technologies in national large-scale testing, as well as different controversies

about this subject. It is believed that such a study is very relevant for Portuguese policymakers and broad stakeholders, as well as for other countries facing similar ongoing processes at present or in the near future.

As such, the objective four is:

Objective 4: To study the extent to which the Covid-19 pandemic context has facilitated the digital transition in Education, and the challenges faced by school communities in the implementation of the DAVE project.

To capture positive and negative impacts during the development and implementation of a computer-based testing initiative (Johnson & Shaw, 2018), different moments of DAVE's implementation are to be documented. In addition, to be able to capture the local educational context, as well as the larger social, political, and economic factors that govern teaching and learning in relation to this new large-scale testing system (Tsagari & Cheng, 2017), a mixed method approach is designed with different stakeholders. In the first part of this sub-study, semi-structured interviews are to be conducted with head teachers from different types of public school in different regions of Portugal. These interviews will allow to elicit views and opinions from the participants (Creswell & Creswell, 2023) regarding this computer-based testing initiative, as well as their concerns about its implications. In the second part of this sub-study, considering the identified head teachers' concerns about DAVE, a semi-structured interview script will be designed, addressed to a representant of IAVE. The interview with a representant of IAVE aims to understand how head teachers' concerns were covered in DAVE's design and implementation. In the third part of this sub-study, to understand to what extent Portuguese teachers use ICT in their practices, their agreement with the reintroduction of external assessments and their agreement with the implementation of DAVE, a survey is to be disseminated among public and private school communities.

3.1.5. Research Framework

Having clarified the objectives of each sub-study, Table 11 summarizes the taxonomy categories for each sub-study, along with their research questions, objectives, data characteristics, and participants. The constructs related to each sub-study are also detailed to provide a clearer understanding of what was targeted in each analysis.

Table 11. Research Design Summary

Agenda	Covid-19 Pandemic Context Student Assessment Policy Student Assessment Practice							
Category	Internal Assessment Practices in ERT	Cancellation of External Assessments	Grade Repetition Beliefs	DAVE's Implementation				
Research Question	Has changing assessment methodologies facilitated teachers' adaptation to ERT?	How was the cancellation of external assessments perceived in schools?	Did the Covid-19 pandemic affected teachers' grade repetition beliefs?	Are schools ready for DAVE's implementation?				
Objective	To investigate whether the enactment of Decree-Law No. 14-G/2020 eased the challenges faced by teachers in their instructional and assessment practices, with a particular focus on teachers of vulnerable groups, such as students with Special Education Needs and Disabilities.	To explore the implications of Decree-Law No. 14-G/2020 on the perception of school head teachers regarding the relationship between internal and external assessment of learning and its impact on school management.	To identify the impacts of the Covid- 19 pandemic on teachers' beliefs towards grade repetition, examining the interaction effect with the MAIA project.	To study the extent to which the Covid-19 pandemic context has facilitated the digital transition in Education, and the challenges faced by school communities in the implementation of the DAVE project.				
Approach	Quantitative	Qualitative	Quantitative	Mix-methods				
Data source	Secondary data (from CNE, 2021a)	Primary data	Secondary data (from Cipriano & Martins, 2021) and Primary data	Primary data				
Data collection methods	1 Survey Questionnaire	32 Semi-structured Interviews	2 Survey Questionnaires	32+1 Semi-structured Interviews & 1 Survey Questionnaire				
Instrument for primary data collection	-	Instrument A	Instrument B	Instrument A, Instrument B, & Instrument C				
Participants	Teachers with coordination functions	Head Teachers	Teachers	Head Teachers, Teachers & IAVE representant				
Data analysis	Mediation and Moderation models	Content analysis	Ordinal logistic regressions or, alternatively, Mann–Whitney U tests	Content analysis & Descriptive statistics				
Constructs	Teachers' adaptation to ERT, Workloads perceptions, Instruction and assessment & Changes to assessment methodologies	Positioning about external assessments cancelation and its reintroduction.	Grade repetition beliefs	Head teachers' concerns, Concerns mitigated by IAVE & Teachers' acceptability of DAVE				

Note that the primary purpose of a taxonomy representation, such as the one in Table 11, is to help users or readers understand and navigate the structure of the subject covered by the taxonomy. As long as it does that effectively, we should not be overly concerned if it does not respect typical standards of a taxonomy frame (Lambe, 2007: 10). Therefore, with the referencing exercise presented in Table 11, the aim is not to reduce the complex reality of student assessment during the Covid-19 period to a traditional compartmentalized taxonomy framework. Instead, it seeks to capture various facets of that reality, thereby strengthening both data collection and the corresponding analysis.

3.2. Sub-study 1: Internal Assessment Practices in ERT

The aim of Sub-study 1 is to analyse the relationships between teachers' adaptation to ERT, their perceptions of workload in ERT, instruction and assessment in ERT, and the effect of changing assessment methodologies. To analyse these relationships, a classic deductive process model was adopted, as depicted in Figure 8.





Source: Adapted from Bryman, 2012: 24

This process, illustrated in Figure 8, invokes what is known in theory about teachers' adaptation to ERT, perceptions of workloads in ERT, instruction and assessment in ERT, as well as the effects of changing assessment methodologies. From the existing literature, hypotheses are deduced, and embedded within these hypotheses are concepts that are translated into researchable entities. These hypotheses are then subjected to empirical scrutiny, i.e., they are tested with empirical data. Finally, the last step involves moving in the opposite direction from deduction to induction, as inferences are drawn regarding the implications of the findings for the theory that initiated the sub-study (Bryman,

2012: 24). To this end, conditional process analysis (with mediation and moderation models) is wellsuited for testing statistical hypotheses.

3.2.1. Method for Data Analysis

According to Hayes (2022), the goal of mediation analysis is to establish the extent to which some putative causal variable, X, influences some outcome, Y, through one or more *mediator* variables (M). One of the simplest forms of such a model is the one depicted in Figure 9. When trying to establish or test how X exerts its effect on Y frequently postulates a model in which one or more intervening variables M is located causally between X and Y. These intervening variables, often called *mediators*, are conceptualized as the mechanism through which X influences Y. That is, variation in X causes variation in one or more mediators M, which in turn causes variation in Y (Hayes, 2022: 6-7).





Source: Adapted from Hayes 2022:7

When the goal is to reveal the boundary conditions of an association between two variables, a moderation analysis is used, as depicted as in Figure 10. A moderation analysis seeks to determine whether the size or sign of the effect of X on Y "interacts with" a moderator variable or more variables. An association between two variables X and Y is said to be moderated when its size or sign depends on a third variable or set of variables W (Hayes, 2022: 7-8).

Figure 10. Simple moderation model with a single moderator variable W influencing the size of X's effect on Y



Source: Adapted from Hayes 2022: 8

The model in Figure 10 represents moderator variable *W* influencing the magnitude of the causal effect of *X* on *Y*. As such, moderation is also known as *interaction* (Hayes, 2022: 8).

Conditional process analysis is employed to describe and test the conditional nature of the mechanisms through which a variable influences another. It combines aspects of both mediation and moderation analyses. Mediation analysis will examine the direct and indirect pathways by which an antecedent variable *X* affects a consequent variable *Y* through one or more mediator variables *M*. Moderation analysis will examine how the relationship between *X* and *Y* is influenced by a third variable or set of variables *W*. Therefore, conditional process analysis integrates these approaches to estimate and interpret how the moderation component affects the mediation process, providing a comprehensive view of the conditional nature of both indirect and direct effects within a causal framework (Hayes, 2022: 10-11). For example, a mechanism linking *X* to *Y* can be said to be conditional if the indirect effect of *X* on *Y* through *M* is contingent on a moderator *W* (Hayes, 2022:413-414), as depicted in Figure 11.





Source: Adapted from Hayes 2022:414

Mediation and moderation analyses are, nowadays, among the most widely used statistical methods in social, behavioural, and health sciences, as well as in business and medical research. These methods are highly regarded for their capacity to examine complex relationships and interactions, making them essential for understanding intricate phenomena and conducting robust analyses (Hayes, 2022: ix). As such, these methods are used in sub-study 1.

3.2.2. Database

To test the relationships between teachers' adaptation to ERT, workloads perceptions in ERT, instruction and assessment in ERT and the effect of changing assessment methodologies, a secondary database will be used. Secondary data analysis refers to the process of using existing data that has already been collected by someone else. Such process offers numerous benefits, including access to high-quality data without the costs and time associated with primary data collection. Additionally, when large samples are used, it allows for subgroup analysis. However, there are also some limitations, such as a lack of familiarity with the data, data complexity, no control over data collection procedures, data quality, and the absence of key variables relevant to the research interests (Bryman, 2012; Vartanian, 2011) that should be considered.

For this sub-study 1, the database created by the CNE – *Conselho Nacional de Educação* [National Council of Education] to support the study *"Educação em tempo de Pandemia: Problemas, respostas e desafios das escolas"* [Education in Pandemic Times: Problems, Responses, and Challenges of Schools] (CNE, 2021a) will be used. The CNE's study aimed to understand how school closures affected teachers, students, and families (CNE, 2021a: 15). To this end, two survey questionnaires were applied during the month of July 2020, after the first ERT experience. One questionnaire was addressed to school head teachers, and the other was addressed to teachers with coordination roles.

The first questionnaire addressed to school head teachers, sought to collect information about the educational contexts. It is important to mention that head teachers, generally, do not have a teaching component at their schools. The second questionnaire, with different questions from the first one, was directed at teachers performing coordination roles in public schools, who also engage in teaching activities. Therefore, the questionnaire included questions that sought the privileged perceptions of these teachers regarding the problems, difficulties, and strategies for dealing with ERT, as well as their perspective as teachers who carried out their duties during that period. For the purposes of this sub-study 1, only the responses from teachers with coordination roles were used.

Regarding the quality of the data from this database, it is important to note that the CNE is a highly reputable organization in Portugal, with studies conducted by highly qualified professionals and academics. Therefore, the reliability, validity, and credibility of the procedures used to collect the
database are assured. Thus, the use of the CNE database for this sub-study offers more benefits than limitations, as the data was collected at the final stage of the first ERT period from a very large sample of Portuguese teachers. Since the data was collected precisely at the end of the first ERT period, the responses given by teachers will provide a more reliable and accurate understanding of the relationships between teachers' adaptation to ERT, their perceptions of workloads, instruction and assessment in ERT, and the impact of changing assessment methodologies. Moreover, given the large sample size, the database will allow testing the same relationships within the subgroup of special education teachers. It is also important to note that the research hypotheses presented in this substudy were not formulated or analysed in the CNE's original study. Thus, this sub-study 1 provides new insights into their data, which was analysed under a research confidentiality agreement.

On the CNE survey to teachers with coordination functions, questions addressed issues related to teachers' roles as coordinators and their experiences with ERT. Among other topics, they covered their own adaptation to ERT, workload perceptions, and instruction and assessment practices. Most of these questions were closed-ended, with single or multiple-choice options. Although, some open-ended questions were also included to enrich the results and address the limitations of survey questionnaires (CNE, 2021a, p. 52). This study, conducted by CNE, is published in Portuguese and it is available on the CNE website³, as well as the questionnaire and main statistics collected through the survey.

3.2.3. Procedures for Data Collection

Teachers' survey for the study conducted by CNE was applied online, preceded by a pre-test, in July 2020. Its dissemination was done through an e-mail sent to all school head teachers at a national level, including autonomous regions of Azores and Madeira. The survey was then forwarded by school head teachers to teachers with coordination functions at their schools. All answers are anonymous, and it is not possible to relate a teacher to another teacher or school (CNE, 2021a: 52).

The sample was formed through a non-probabilistic sampling process (CNE, 2021a: 54). Still, note that for the purposes of defining randomness for survey sampling it is sufficient to assert that randomness means a lack of pattern or predictability (Stopher, 2012: 68-70). Even if the sample was obtained through a non-probabilistic process, it should be noted that there is no pattern or predictability on the sample.

It is estimated that the 4 338 valid answers to the survey correspond to a 29% rate of teachers with coordinating functions at a national level (CNE, 2021a: 54).

³ See <u>https://www.cnedu.pt/pt/noticias/cne/1673-educacao-em-tempo-de-pandemia-problemas-respostas-e-</u> desafios-das-escolas.

3.2.4. Participants

Preschool teachers with coordination functions at schools also participated in CNE's survey. It is important to note that, in Portugal, preschool, for children aged from 3 to 5 years old, does not take part of compulsory schooling system. Moreover, Portuguese preschools have specific orientations regarding instruction, curriculum, and assessment (see OCEPE by DGE, 2023). Therefore, for this given sub-study about assessment of learning in compulsory school, preschool teachers that participated on the survey were excluded from the CNE database, leading to a new sample of 3 932 teachers. Figure 12 shows the distribution of these 3 932 teachers by NUT III.



Figure 12. Teachers' distribution by NUT III



Source: CNE teachers' survey, 2020

Note as well that coordination functions at schools are usually performed by permanent school board teachers. Consequently, only a small number of young teachers under 40 years old perform these functions in schools, which explains the age bias in the sample (χ^2 (2, N = 3,830) = 689.62, p < .001). In fact, teachers younger than 36 years old performing coordination functions are considered statistical outliers, as shown in Figure 13.





	Sample		Populatio	on
	n	%	Ν	%
Sex				
Female	2 991	76.3	91 466	75.5
Male	928	23.6	29 660	24.5
Age				
< 39 years old	83	2.2	14 396	11.9
Between 40 – 49 years old	1 026	26.8	44 355	36.6
≥ 50 years old	2 721	71.0	62 375	51.5

Table 12. Teachers with coordination functions by sex and age

Source: CNE teachers' survey, 2020; DGEEC, 2021

Source: CNE teachers' survey, 2020

Accordingly, Table 12 shows that coordination functions are mostly performed by mid-career teachers, where 71.0% are above 50 years old, with a mean of 53.3 years old (*SD* = 6.7 years old). Regarding sex, the sample fits the population (χ^2 (1, N = 3 919) = 1.38, *p* = .240). Women teachers in the sample are prevalent (76.3%), as they are in the population (Table 12).

In order to provide an understanding regarding teachers' subjects areas, teachers' recruitment group (qualitative variable containing 33 different recruitment groups) was recoded into 6 categories to better capture its core domains, as depicted in Figure 14.



Figure 14. Teachers (%) by curricular area

Source: CNE teachers' survey, 2020

3.3. Sub-study 2: Cancellation of External Assessments

The aim of Sub-study 2 is to explore the perceptions of school head teachers regarding the cancellation of external assessments during the Covid-19 pandemic, its implications for the relationship between internal and external assessments of learning, and its impact on school functioning. To achieve this, a qualitative approach was chosen to collect primary data through a series of in-depth, semi-structured interviews. In qualitative research, the focus is on understanding the meanings that participants attach to a problem or issue, rather than the meanings researchers bring to the study or those found in the literature (Creswell & Creswell, 2023: 225). Therefore, the purpose of an in-depth interview is not to obtain clear answers or to test hypotheses. At the core of in-depth interviewing is the importance to

understand the lived experiences of others and the meanings they derive from those experiences (Seidman, 2006: 9). Capturing head teachers' understandings of their lived experiences regarding the cancellation of external assessments during the Covid-19 pandemic is, therefore, the focus of Substudy 2. To accomplish this, it is necessary to collect primary data, which refers to data gathered for the first time and original in nature (Kothari, 2004: 95). In studies using primary data, researchers are responsible for both collecting and analysing the data (Bryman, 2012). The interview protocol for head teachers' interviews is referred to as Instrument A, which can be found in Appendix A.

3.3.1. Interviewing

According to Kothari (2004: 98-99) the interview method of collecting qualitative primary data involves the presentation of oral-verbal stimuli and receiving responses in the form of oral-verbal replies. When considering the purposes of Sub-study 2, the major advantages of interviews are:

- 1. More information with greater depth can be obtained.
- 2. The possibility to overcome any resistance from the head teachers.
- 3. Greater flexibility to restructure questions if needed.
- 4. Sample can be controlled more effectively, as there is no difficulty with non-response during the interview with head teachers.
- 5. Greater control on which head teachers will answer the questions.
- 6. Possibility to collect supplementary information about the schools and head teachers contextual background, which is of great value in interpreting results (Kothari, 2004: 98-99).

Still, it is recognized that interviewing in Sub-study 2 also has important weaknesses, such as:

- 1. Expensive method, considering that a large and widely spread geographical sample is involved.
- 2. Time-consuming for data collection and analysis.
- 3. Certain types of schools' head teachers may not be easily approachable.
- 4. The presence of the interviewer may over-stimulate or inhibit the head teachers responses.
- 5. An effective interview requires proper interaction with respondents to facilitate free and frank responses, which can be a difficult requirement when head teachers talk about their schools.
- 6. There remains the possibility of bias from both the interviewer and the respondents.
- 7. Interviewing may also introduce systematic errors at times (Kothari, 2004: 98-99).

To overcome some of these limitations, it was decided to conduct interviews via videoconference call. This approach allows for visual contact and interaction with the interviewee, while also facilitating nationwide coverage and cost control. Hence, the use of webcam makes the online interview similar to a telephone interview in that it is mediated by technology, but also similar to an in-person interview, as those involved in the exchange can see each other (Bryman, 2012: 669). Additionally, participants may find online interviews less stressful and more convenient due to the comfort and familiarity of their own environment (Salmons, 2022: 47).

3.3.2. Instrument A Development – Interview Protocol

In qualitative interviews, interviews generally involve few open-ended questions and are intended to elicit views and opinions from the participants (Creswell & Creswell, 2023: 230). Structured interviews involve the use of a set of predetermined questions and highly standardized recording techniques. In contrast, unstructured interviews are characterized by a flexible approach to questioning, without a system of predetermined questions or standardized recording techniques (Kothari, 2004: 97-98). Semi-structured interviews aim to combine the strengths of these two techniques by using a set of predetermined questions while maintaining flexibility in the approach to questioning. These questions are often broader in scope compared to those in a structured interview. Additionally, the interviewer generally has the freedom to ask follow-up questions based on responses (Bryman, 2012: 212).

The interview protocol comprises several important components, including basic information about the interview, an introduction, the interview content questions with probes, and closing instructions. In the introduction section, the interviewer introduces himself, discusses the purpose of the study, explains the general structure of the interview, and asks the interviewee if they have any questions before beginning. The content questions are the research sub-questions of the study and essentially break down the central phenomenon into its parts. These content questions include probes, which are reminders for the researcher to seek more information or ask for explanations of ideas. In the closing instructions, the interviewer thanks the interviewee for their time and responds to any final questions. Debriefing reassures the interviewee about the confidentiality, while offering a summary of the final study (Creswell & Creswell, 2023: 235-236).

Instrument A – interview protocol (available in Appendix A), is organized into 3 blocks of questions:

- 1. Internal assessments in the Covid-19 Pandemic
- 2. External assessments in the Covid-19 Pandemic
- 3. The future of student assessment

As Decree-Law No. 14-G/2020 placed the responsibility on schools to handle student assessment during the COVID-19 pandemic, Block 1 contains four content questions designed to understand the context and internal assessment practices that each school adopted during the Covid-19 pandemic. Block 2 contains five content questions regarding the purposes and the cancellation of external assessments. Block 3 contains six content questions about the future of student assessment and the relationship between internal and external assessments. The interview script was developed based on literature regarding the Covid-19 pandemic in Portugal (such as Flores & Gago, 2020; CNE, 2021a) and based on the content of an exploratory interview with the board of a school in the Lisbon Area.

3.3.3. Sampling and Data Analysis

Given the purposes to capture local educational contexts, as well as the larger social and administrative factors (Tsagari & Cheng, 2017), a sampling process close to the theoretical sampling process (Strauss & Corbin 2015: 337-340) was implemented, considering different i) School cluster type, ii) School administrative and governance characteristics, and iii) Regions by Nomenclature of Territorial Units II, level 2 (NUTS II). The process of data collection was controlled by the emerging information, and the achievement of theoretical saturation was a criterion for deciding when to cease collecting new data (Bryman, 2012: 418-420), as depicted in Figure 15.





Source: Adapted from Bryman, 2012: 420

As potential participants for this sub-study 2, it was decided to exclude head teachers from the autonomic regions of Azores and Madeira due to their autonomic administration characteristics regarding the Portuguese central government. In mainland Portugal, head teachers from private schools were also excluded due to their private administration characteristics in relation to the Ministry of Education policy. In addition, from the list of all public school clusters and non-clustered schools in mainland Portugal (N=812; Source: DGAE, 2019), head teachers from professional or artistic schools were excluded. This is because these types of schools do not require external assessment tests to conclude compulsory education, relying instead on vocational or artistic aptitude tests conducted at the school level to conclude compulsory education.

Then, to gather participants for the interviews, an e-mail was sent to all the other head teachers in mainland Portugal (n=781) promoting this study. As head teachers responded to the sent e-mail to voluntarily participate in this study, interviews were conducted via a videoconference call, from mid-October 2022 until the end of February 2023. As interviews were carried out, recorded and transcribed, a content analysis was performed (Bardin, 2014) through an open coding process.

The purpose of content analysis is to reduce the amount of data a researcher has to work with by delineating concepts to stand for data (Strauss & Corbin, 2015: 194). In interview analysis, coding means that a section of interview text is selected and linked to a label (a code, a category). From a formal point of view, a coded text segment always consists of two elements: the text passage and the assigned category (code). Such process requires the existence of coding rules on how should the scope of a coded segment be determined and how should a repetitive statements be dealt with (Rädiker & Kuckartz, 2020: 54-55). An open coding process is an exploratory coding process for concepts identification. It breaks data apart, delineating concepts to stand for interpreted meaning of raw data (Strauss & Corbin, 2015: 531).

Note that hand coding is a laborious and time-consuming process, even for data from a few individuals. Thus, qualitative software programs have become popular (Creswell & Creswell, 2023: 237), and to analyse and report data within this sub-study 2, MAXQDA 2022 software is used. This software, like other Qualitative Data Analysis programs (QDA), helps researchers organize, sort, and search for information in text. Through the assignment of codes, this computer process is faster and more efficient than hand coding. Additionally, researchers can quickly locate text segments coded the same way and determine whether participants respond to a code's idea in similar or different ways, helping them recognize when no new information is emerging. Beyond this, the computer program can facilitate relating different codes for report. Note that these softwares does not analyse the data; it is the researcher's responsibility to conduct the analysis (Creswell & Creswell, 2023: 238).





Figure 16 shows the total number of analytical categories created and used in MAXQDA (n=265), as the number of performed and analysed interviews grew. Considering that no new nor relevant data was emerging regarding analytical categories on the latest analysed interviews; and considering that the existing categories were already well developed in terms of their properties and dimensions, demonstrating variation with well-established relationships among the categories (Bryman, 2012: 241); it was considered that the theoretical saturation has been achieved after conducting 32 interviews, with a total of 24 hours and 46 minutes length.

3.3.4. Participants

Table 13 shows how the characteristics of each school cluster type relate with other characteristics (Administrative and Governance characteristics, and Region by NUT II). Head teachers from Basic + Secondary school clusters, with regular administration in the Centre Region of Portugal are predominant in the sample.

Source: Primary data, 2023

Interview	Scho	ol Cluster	Туре	Adm governa	ninistrative nce charae	e and cteristics	Region	by Nomen	clature of	Territoria	l Units II
Number	Basic only	Basic + Secondary	Secondary only	Regular	TEIP	Autonomy contract	North	Centre	Lisbon M. Area	Alentejo	Algarve
01		Х			Х						Х
02		Х				Х		Х			
03		Х		Х						Х	
04		Х		Х						Х	
05		Х		Х					Х		
06	Х		1		I	Х	Х				
07		Х				х			Х		
08	Х				Х					Х	
09		Х		Х				Х			
10			Х	Х				Х			
11		Х			I	х	Х				
12		Х		Х				Х			
13	Х				Х					Х	
14	Х			Х				Х			
15		Х			1	Х		Х			
16		Х		Х			1	Х			
17		Х			Х				Х		
18	Х			Х				Х			
19		Х		Х					Х		
20		Х		Х			Х				
21	Х				I	х	Х				
22		Х		Х					Х		
23	Х			Х							х
24		Х		Х			Х				
25			х	Х					Х		
26	Х			Х				Х			
27		Х		Х				Х			
28			х	Х							х
29		Х		Х							Х
30			х	Х			Х				
31	Х			Х						Х	
32		Х		Х						Х	
TOTAL	9	19	4	22	4	6	6	10	6	6	4

 Table 13. School Characteristics by Cluster Type, Administrative and governance characteristics, and Region (n)

Source: Primary data, 2023

Table 14 shows the characteristics of the schools of the 32 interviewed head teachers, and how representative they are (in percentage) comparing to school population in mainland Portugal. While the percentage of school cluster types and administrative and governance characteristics are very similar between the sample and the population, there is an underrepresentation of schools from the Lisbon Metropolitan Area and the North region, and an overrepresentation of schools in the Centre, Alentejo, and Algarve regions.

		Sample	ł	Population
	n	%	Ν	%
Type of School Cluster				
Basic Education Only	9	28.1	204	25.8
Basic + Secondary Education	19	59.4	519	65.5
Secondary Education Only	4	12.5	69	8.7
Administrative and Governance Characteristics				
Regular	22	68.8	495	62.5
TEIP – [Educational Territories of Priority Intervention]	4	12.5	135 ^{a)}	17.0 ^{a)}
Autonomy contract	6	18.8	208 ^{a)}	26.3 ^{a)}
Region by Nomenclature of Territorial Units II				
North	6	18.8	282	35.6
Centre	10	31.3	145	18.3
Lisbon Metropolitan Area	6	18.8	258	32.6
Alentejo	6	18.8	67	8.5
Algarve	4	12.5	40	5.0

Table 14. School Characteristics by Cluster Type, Administrative and Governance Characteristics, and Region (n, N)

Notes: a) In the population, 46 schools (5.8%) are TEIP schools with an autonomy contract. Sources: DGEEC, 2023; DGAE, 2019; GesEdu, 2023; Primary data, 2023 (Own calculations)

Additionally, to provide a geographical understanding, Figure 17 illustrates the distribution of head teachers' schools throughout the Portuguese mainland territory by NUTS II – Nomenclature of Territorial Units for Statistics, Level 2.

Figure 17. Head Teachers and schools' distribution (n) by NUT II

10

Interviews (n)



Source: Primary data, 2023

3.4. Sub-study 3: Grade Repetition Beliefs

Beliefs are not directly observable and can only be inferred from responses or indicators of their practice (Eagly & Chaiken, 1993). To understand how the Covid-19 pandemic affected Portuguese teachers' grade repetition practices, institutional indicators were analysed in section 2.3.3. To understand how the Covid-19 pandemic affected Portuguese teachers' grade repetition beliefs, a quantitative approach will be implemented to collect primary data and compare responses given by teachers in two survey questionnaires, using two independent samples from different time periods: Sample A, collected before the Covid-19 pandemic in April 2019, and Sample B, collected in the post-pandemic period in October 2023.

3.4.1. Sample A

The database created by Cipriano & Martins (2021) to support the study 'Beliefs on assessment and grade repetition among teachers in Portugal' was used to form sample A. This exploratory study, based on a survey questionnaire administered in April 2019, aimed to understand the conceptions and beliefs of Portuguese teachers about student assessment, and whether their beliefs are in line with the guidelines on national education policy. Moreover, it aimed to understand why Portuguese teachers use grade repetition as a common practice when it is not recommended. From the Cipriano & Martins (2021) database, four items regarding teachers' grade repetition beliefs and practice were selected:

- Item 1.1. Grade repetition should only occur by the end of a cycle;
- Item 1.2. For some students, grade repetition is inevitable;
- Item 1.3. Grade repetition is beneficial for students;
- Item 1.4. I use grade repetition as an educational measure.

On the four items, a ten-point Likert scale was used, where 1 means "Totally disagree" and 10 means "Totally agree".

3.4.2. Sample B

To collect Sample B, a survey questionnaire (Instrument B) was designed to gather primary data. In addition to the close-ended questions, participants were asked to gauge the extent to which they hold specific attitudes or perspectives on certain research topics through statements in the questionnaire.

This technique, developed by Rensis Likert with the creation of the Likert scale, allows respondents to indicate their level of agreement or disagreement (Babbie, 2013: 231).

The survey questionnaire to collect Sample B was prepared to be conducted using the same procedures as those used for Sample A, replicating the previously selected four items from Sample A. The selection of items previously used in other studies allows us to compare data from two time periods with two independent samples (Sudman & Bradburn, 1982). Note that in the study conducted by Cipriano & Martins (2021), a very large Likert scale (with 10-points) was used to allow the analysis of item responses with quantitative approaches. However, throughout the analysis, researchers found that the assumption of equal variances within each group of teachers was not met, preventing the application of ANOVA tests. Therefore, researchers only used non-parametric tests in their analysis. As such, it does not make sense to use such a large Likert scale for Sample B collection, and a five-point Likert scale will be used, allowing responses to be given in the intended dimensions with verbal points (Sudman & Bradburn, 1982), on a scale from 1 (Totally disagree) to 5 (Totally agree). To enable comparison between samples on data analysis, the 10-point Likert scale from Sample A will be recoded as a 5-point scale. Furthermore, on Instrument B it will be asked whether teachers have participated in the MAIA training project to assess the effect of participation with these four items. Additionally, to understand teachers beliefs towards grade repetition, questions about who is most responsible for a student's success or grade repetition, as well as about post-pandemic assessment practices are asked.

3.4.3. Instrument B Development – Survey Questionnaire

According to Bryman (2012: 232-237), one of the main instruments for gathering data using a social survey design is the self-administered questionnaire, which is completed by respondents themselves. Sill according to Bryman (2012: 676-677), web based self-administered questionnaires are becoming more popular and operate by inviting prospective respondents to visit a website at which the questionnaire can be found and completed online. These surveys offer many advantages such as:

- Low cost,
- Fast response;
- Absence of interviewer effects;
- Convenience for respondent;
- No constraints in terms of geographical coverage;
- Few unanswered questions as it is controlled automatically;
- Good data accuracy, as data entry is automated.

Nevertheless, self-administered web questionnaires also have disadvantages such as:

- Low response rate;
- Restricted to online populations;
- Requires motivation;
- Multiple replies (Bryman, 2012: 676-677).

Despite the existing disadvantages, considering the purposes of this investigation as well as the fact that Sample A was collected using a web based self-administered questionnaire, the same approach will be replicated for Sample B collection. Consequently, the questionnaire is designed to be conducted online using Qualtrics software. Note that Instrument B will also be used to collect data for Sub-Study 4. Therefore, in developing the questions and statements for Instrument B, information gathered through Instrument A was also considered. Additionally, some other questions from Cipriano & Martins (2021) study were also replicated, as well as some questions and statements from CNE (2021a), and Rodrigues et al. (2022) studies. Furthermore, 10 additional items were included to capture information on age, sex, region, qualification, and other professional backgrounds to characterize the sample.

The Instrument B questionnaire contains both statements and closed-ended questions. Incorporating both statements and questions in a questionnaire allows for greater flexibility in item design and can make the survey more engaging. Additionally, closed-ended questions are particularly popular in survey research because they ensure more consistent responses and are easier to analyse than open-ended ones (Babbie, 2013: 231). By the end of the questionnaire, it was created an open-ended question to allow respondents to elaborate. The open-ended question (sometimes called open question) is a question in which possible answers are not suggested, and the respondent answers in his or her own words (Popping, 2015).

Before its implementation, the survey questionnaire was pre-tested by two academic researchers to identify flaws in its design and technical aspects, and by three teachers to identify issues with the content. Instrument B for data collection can be found in Appendix B.

3.4.4. Sampling Processes for Data Collection

Data supporting the Cipriano & Martins (2021) study, with n = 1 844 teachers, was collected through a survey questionnaire sent by email to all head teachers of public and major private schools in mainland Portugal, and then forwarded and distributed among teachers at their school communities. When sampling for quantitative approaches where the goal is to measure a large population through a sample that represents the entire population, representativeness becomes essential (Stopher, 2012). However, web-based surveys have significant issues and limitations, particularly the fact that not everyone in a given population has internet access, or not everyone has the technical ability to handle questionnaires in either email or web formats (Bryman, 2012: 673). Still, note that teachers population is a qualified one, and all teachers have access to the internet, at least, at schools.

According to Stopher (2012: 68-69), in statistical terms, representativeness can be defined in the following way:

- 1. Sample means are statistically no different from population means;
- 2. Sample variances are statistically no different from population variances; and
- 3. Sample covariances are statistically no different from population covariances.

Representativeness can generally be achieved when the probability of any element in the population being included in or excluded from the sample is known. The only well-recognized method for ensuring known probabilities of inclusion or exclusion is through random sampling (Stopher, 2012: 68-69). However, this process can be very challenging for a single researcher to manage when dealing with large populations, such as the national population of teachers, without a complete nominal listing of the entire group.

Nevertheless, note that one of the definitions of randomness has to do with a lack of pattern or predictability (Stopher, 2012: 69). With this in mind, to form sample B with the objective of ensuring a good quality sample close to the procedures of a probabilistic sample while being consistent with previous sampling processes used in sample A, the designed survey questionnaire (Instrument B) was disseminated through an email sent to all head teachers at all public schools in Portugal, including the autonomous regions of Azores and Madeira. In addition, an email with the survey link was also sent to all head teachers at private schools that performed national secondary education exams during the 2022/2023 school year. Emails were sent to head teachers during the last week of September 2023, asking them to forward the survey to all other teachers in their schools and to voluntarily participate in the study throughout October 2023. In mid-October, the solicitation was reinforced. These procedures, which included all public schools and major private schools, allowed for the collection of a very large sample with no patterns or predictability regarding which teachers, from which schools, would respond. Additionally, it is important to note that increasing the sample size generally enhances the sample's representativeness (Stopher, 2012: 65). Therefore, having a very large high-quality sample is a mean to increase representativeness.

3.4.5. Participants

The procedures used to collect Sample B involved 3 418 engaged participants with 2 673 validated responses, gathering data from all regions of Portugal. Figure 18 shows the distribution of the 2 673 teachers in Sample B by NUTS III – Nomenclature of Territorial Units for Statistics, Level 3.



Figure 18. Sample B Teachers' distribution by NUT III

Source: Primary data (2023)

As the two samples under study (Sample A – collected before the Covid-19 pandemic and Sample B – collected in the post-pandemic period) were formed through non probabilistic processes, by performing the Chi-squared goodness-of-fit test, it is possible to observe sampling bias regarding sex,

age, qualification, school sector and teaching level. However, it should be once again noted that there is no pattern or predictability in these two samples, and it is not possible to relate a teacher to another teacher or school. Additionally, these sampling procedures allowed the creation of two very large samples which, in general, have proportions (in %) very similar within the samples and the population, as shown in Table 15.

		Sample A		Samplo P	Population a)		
		(n = 1.844)		(n = 2.673)	(N = 124.066)		
			(<u>_</u> 0,0)		N	" 121000) %	
Sov		Valia /		Valia /0		70	
Sex .	4 2 2 4	72.2	2 050	70.0		75.0	
Female	1 3 3 1	72.2	2 058	/8.6	93 999	/5.8	
Male	513	27.8	561	21.4	30 067	24.2	
Age							
≤ 39 years old	166	9.0	160	6.0	12 064	9.7	
Between 40 – 49 years old	658	35.7	923	34.5	44 140	35.6	
≥ 50 years old	1 020	55.3	1 590	59.5	67 862	54.7	
Qualification							
ISCED 5	45	2.4	31	1.2	6 368	5.1	
ISCED 6	1 410	76.5	2 036	76.2	99 753	80.4	
ISCED 7 or 8	389	21.1	606	22.7	17 945	14.5	
School sector							
Public	1 778	96.4	2 503	93.6	112 160	90.4	
Private	66	3.6	170	6.4	11 906	9.6	
Teaching level ^{b)}							
1 st cycle (grades 1-4)	393	22.4	508	20.2	28 608	23.1	
2 nd cycle (grades 5-6)	296	16.9	429	17.0	21 613	17.4	
3 rd cycle / Secondary (grades 7-12)	1 064	60.7	1 581	62.8	50 221	59.5	

Table 15. Teachers from the compulsory school system by sex, age, qualification, school sector and teaching level

Notes: a) The population data refers to the 2021/2022 school year in mainland Portugal, excluding Madeira and Azores. b) Special Education' and Religious Education' teachers were not considered as they teach at one or more levels. Sources: Cipriano & Martins, 2021; DGEEC, 2023; Primary data, 2023

Both in the samples and in the population, Table 15 shows that female teachers are prevalent. Furthermore, most teachers are older than 50 years, with a bachelor's degree or equivalent tertiary education level (ISCED 6) as a qualification. Hence, the great majority of teachers in Portugal work in the public sector, and ca. 60% of teachers work with students in lower and/or upper secondary education (grades 7-12).

Note as well that in sample B, 1 703 teachers (63.7%) reported that they did not participate in the MAIA project (defining subsample B_0), and 970 teachers (36.3%) reported having participated in the MAIA training project (defining subsample B_1). This will allow to study the effect of participation in the MAIA project on teachers' grade repetition beliefs.

3.4.6. Methods for Data Analysis

Considering that sampling bias exists, to compare the distributions of the four ordinal variables in the two independent samples A and B while controlling for sex, age, qualification, school sector and teaching level as covariates, ordinal logistic regressions are performed with these variables as covariates, using complementary log-log, negative log-log, and probit functions according to class distributions (Marôco, 2021). Ordinal logistic regression is a statistical technique used for modelling ordinal variables. However, to use this technique, there is data assumptions, and it is required that data fits the models (Harrell, 2015). As in all models the data does not fit the model well and the slope homogeneity is rejected for all the items (item 1.1. χ^2_{PL} (27) = 131.807, p < .001; item 1.2. χ^2_{PL} (27) = 157.858, p < .001; item 1.3. χ^2_{PL} (27) = 249.490, p < .001; item 1.4. χ^2_{PL} (27) = 105.211, p < .001), logistic regressions are only used as exploratory analysis.

Alternatively, Mann–Whitney U tests are performed to compare the distributions of the two independent samples A and B and to determine whether there is a significant difference between the distributions on the two samples. The Mann-Whitney U test is a nonparametric test that may be used when the data assumptions cannot be met. The test uses the ranks of the values. The use of ranks only requires that the data is measured at the ordinal level. The ultimate purpose of the Mann-Whitney U test is to search for statistical evidence that the samples are significantly different (Aldrich & Cunningham, 2016: 139-140).

To understand the effect of participation on the MAIA project, the very same procedures are then followed to compare the distributions between sample A *vs* subsample B₀, sample A *vs* subsample B₁ and subsample B₀ *vs* subsample B₁. Further, to understand the effects of sex, age, qualification, school sector and teaching level, all previous tests are also performed selecting the categories of these five variables as subgroups to be tested.

The Mann-Whitney U test is also used to understand teachers' practices in the post-pandemic period, considering whether or not they participated in the MAIA project.

3.5. Sub-study 4: DAVE's Implementation

The aim of Sub-study 4 is to examine the extent to which the Covid-19 pandemic has facilitated the digital transition in education and the challenges faced by school communities in implementing the DAVE project. To achieve this, a three-stage mixed-methods approach, involving three different stakeholders, is designed. As with typical mixed-methods approaches, Sub-study 4 uses qualitative data to provide context for interpreting broad quantitative findings. Additionally, the inclusion of different stakeholders in mixed-methods research is believed to enhance its credibility among audiences, as it captures multiple perspectives. Therefore, this mixed-methods approach, with qualitative and quantitative data from different stakeholders, is often preferred as it is seen as more likely to produce findings with practical value (Bryman, 2012: 645-650).

To this end, first, within Instrument A (interviews with 32 head teachers), participants are asked whether their school is ready for the implementation of DAVE, and why. This qualitative approach allows for identifying the challenges and concerns (if any) that schools face in implementing DAVE, through the elicited views and opinions of the participants (Creswell & Creswell, 2023: 230). As mentioned earlier, the interview protocol for head teachers can be found in Appendix A.

The second stage, also qualitative in nature, involves the development of a new data collection instrument (Instrument C, which can be found in Appendix C). This instrument is created based on the information that emerged from the interviews with head teachers, and it is addressed to a representant of IAVE, where the president of this institute decided to respond for the organization. The purpose of this interview is to understand the extent to which the concerns raised by the school head teachers are taken into account in the design and implementation of DAVE.

In the third stage of sub-study 4, which is a quantitative approach, the aim is to understand to what extent Portuguese teachers use ICT in their classroom practices, with a particular focus in the use of ICT in classroom testing. Furthermore, it is sought to gauge their agreement with the reintroduction of external assessments in the Portuguese educational system and their standing regarding the implementation of DAVE. These questions are addressed through Instrument B (the survey questionnaire for teachers), which can be found in Appendix B.

3.5.1. Instrument C Development – Interview Protocol

The protocol for the interview to the president of IAVE is designed based on the information that emerged from the 32 interviews with head teachers. The concerns raised by head teachers were of two different types:

- 1. The testing system itself, tests constructs, and tests layout;
- 2. Schools' conditions for its implementation.

Considering head teachers concerns, the semi-structured interview protocol contains two blocks of questions. The first block includes questions about the tests' constructs, curriculum limitations, accessibility and universal design of computer-based large-scale tests, students' learning contexts, and alternative measurement devices to large-scale testing. The second block of questions, related to the implementation process of DAVE, contains questions about the scarce information provided by IAVE regarding DAVE's implementation, infrastructure and human resource limitations to implement DAVE, and technical and security issues about DAVE. As referred before, this interview protocol (Instrument C) can be found in Appendix C.

3.5.2. Participants and Data Analysis

As referred to in Section 3.3, the data collected from the 32 semi-structured interviews with head teachers was analysed using MAXQDA software. A part of this data, concerning the implementation of DAVE, is used in the first stage of this sub-study 4. Since this stage of sub-study 4 concerns to the same set of interviews performed in sub-study 2, for information regarding the participants characteristics and data analysis procedures, please see sub-section 3.3.4. and sub-section 3.3.3., correspondingly.

In the second stage of sub-study 4, the interview with the president of IAVE was conducted in person, at IAVE's headquarter in Lisbon, in July 2023, after the first implementation of PA in digital format, during the 2022/2023 school year. With the participant consent, the interview was recorded and later transcribed to allow content analysis (Bardin, 2014). Then, the qualitative data was also analysed with MAXQDA software (Rädiker & Kuckartz, 2020), taking into account the concerns (categories) identified from the head teachers interviews. The procedures used during the coding process were the same as in the previous 32 interviews with head teachers, as the aim of this analysis was to understand how head teachers' concerns were addressed in the design and implementation of DAVE.

In the third stage of sub-study 4, quantitative data collected through Instrument B was analysed using SPSS software. Descriptive statistics are used to report the findings in Chapter 7. Note that this database refers to the same database as sub-study 3. Therefore, for information on participants characteristics, please see section 3.4.5, regarding the characteristics of Sample B.

3.6. Ethical Issues

Ethics in research is not a detail or a set of boring procedures. Instead, it is at the heart of the entire research process. According to Bryman (2012: 135), there are four ethical principles that must be considered:

- 1. Whether there is harm to participants;
- 2. Whether there is a lack of informed consent;
- 3. Whether there is an invasion of privacy;
- 4. Whether *deception* is involved.

Still according to Bryman (2012: 138), research that is likely to harm participants is regarded by most people as unacceptable. When potential risks to participants are anticipated, informed consent must be obtained. Although this research project about student assessment during the Covid-19 pandemic appears to pose no risks to participants, nor involves any invasion of privacy or deception, informed consent was still obtained from all participants.

Inquiries involving human subjects should be based as far as practicable on the freely given informed consent of subjects (...). In voluntary inquiries, subjects should not be under the impression that they are required to participate. They should be aware of their entitlement to refuse at any stage for whatever reason and to withdraw data just supplied. Information that would be likely to affect a subject's willingness to participate should not be deliberately withheld, since this would remove from subjects an important means of protecting their own interests. (...) As far as possible, participation in sociological research should be based on the freely given informed consent of those studied. This implies a responsibility on the sociologist to explain as fully as possible, and in terms meaningful to participants, what the research is about, who is undertaking and financing it, why it is being undertaken, and how it is to be promoted. (Bryman, 2012: 138-139).

Addressing ethical issues is also an important dimension when research is conducted online (Salmons, 2022). Therefore, the informed consent of each head teacher interview was obtained by having them sign informed consent forms with digital signature. This form was prepared according to lscte' template and provided information about the study, the nature of each participant's involvement, how data would be treated (including personal data), and who would be responsible for

handling that data. The informed consent also detailed confidentiality and voluntary participation, how data would be stored, for how long, and the purposes for which the data would be used. Additionally, it provided the researcher's contact information and the contact information for Iscte's Data Protection Officer. The informed consent form used to inform head teachers before their interview can be found in Appendix D.

The form for informed consent and procedures for conducting the interviews were approved by Iscte's Ethics Committee, as also detailed in Appendix D. This informed consent form and procedures were then fine-tuned for the interview with the president of IAVE, a public figure who agreed to have their responses published (Babbie, 2013: 36), which consent was also collected through a signed form. For the online survey questionnaire with teachers, consent was obtained by their free advancement in the web survey.

In addition to ethical obligations that researchers have to subjects, researchers have also ethical obligations to their colleagues in the scientific community. These obligations concern the analysis of data and the way the results are reported (Babbie, 2013: 39). Therefore, in addition to ethical obligations to all participants, Iscte's ethical procedures were followed throughout the research for data analysis, and reporting.

3.7. Chapter Final Considerations

Considering the research design developed in this chapter for each sub-study (Sub-study 1: Internal Assessment Practices in ERT, Sub-study 2: Cancellation of External Assessments, Sub-study 3: Grade Repetition Beliefs, and Sub-study 4: DAVE's Implementation), in the next four chapters, the theoretical framework of each sub-study is further developed, and the research results and findings are reported.

This page intentionally left blank

CHAPTER 4

Assessment of Learning During the 1st Wave of the Covid-19 Pandemic with a Glance at Special Education

Chapter Summary:

With the covid-19 lockdowns, teachers, around the world, have had to adapt quickly to emergency remote teaching, as well as their instruction and assessment practices. This transition to emergency remote teaching has resulted in heavier workloads for teachers. Using data from a questionnaire applied to 3 932 teachers in Portugal, this chapter' research showed that the relationship between teachers own adaptation to emergency remote teaching with perceived workloads was mediated by instruction and assessment, and this indirect effect was moderated by changes made on assessment methodologies. Therefore, changing assessment methodologies during this period was an important strategy for a better adaptation to emergency remote teaching. In addition, when special education teachers were considered, it was observed that changing assessment methodologies did not moderate the relationship between teachers' adaptation with instruction and assessment. These findings allow us to conclude that the difficulties experienced by special education teachers went beyond changing assessment methodologies.

4.1. Hypotheses Formulation

4.1.1. Teachers' Adaptation to ERT and Perceived Workloads

With imposed teleworking during the covid-19 lockdowns in most countries, concerns regarding working from home have arisen; and teleworking conditions became a fertile research field for academics. According to Palumbo (2020), home-based teleworking affects negatively work-life balance of public servants. Employees who remotely worked from home suffer from a greater work-related fatigue. Teleworking from home has side effects on remote workers' fatigue due to intensification and extensification of work efforts (Palumbo, 2020).

During the covid-19 lockdowns, teachers' adaptation to ERT and teleworking from home, implied the realization of several new tasks to be performed by students and teachers (E.g., CNE, 2021a; Flores & Gago, 2020; OECD, 2020b, 2021). Those new tasks at home have induced on teachers the feeling of an increased workload during emergency remote teaching (CNE, 2021a; Klusmann et al, 2023), leading to various consequences in teachers' health conditions, including depression, anxiety, stress, and burnout syndrome (Jerrim et al., 2024; Lizana & Vega-Fernadez, 2021), related to difficulties in managing family and professional roles, in creating the best conditions for teaching, learning and assessment, and in technical aspects linked to software and use of technology (Flores et al. 2023). Therefore, we predicted that:

H1: Teacher's adaptation (TA) is positively related to workloads (WL).

Figure 19. Conceptual model for Hypothesis 1



4.1.2. Instruction and Assessment During ERT (The Mediator)

In March 2020, around the world, the implementation of distance teaching due to the Covid-19 pandemic imposed profound restrictions to teachers' practice with consequences on students' learning (E.g., Flores & Gago, 2020; CNE, 2021a, 2021b; Esteves et al., 2021; Ikeda & Echazarra, 2021; OECD 2020b, 2021) including educational practices and support to students and families with special needs (Simó-Pinatella et al., 2022).

In Portugal, schools' closure due to the Covid-19 pandemic has transferred from face-to-face to ERT approximately 1.5 million students and 135 thousand teachers (Source: DGEEC). Although it was an important first response to the Covid-19 pandemic, distance teaching on this period was, in fact, an unprepared, improvised, and temporary solution. Moreover, it was not performed by professionals with digital and pedagogical skills for distance teaching (OECD, 2020b; CNE, 2021a). Thus, this form of teaching – Emergency Remote Teaching (ERT) – must not be confused with online learning and other forms of intentional distance teaching and distance learning (Hodges et al., 2020).

In addition to the restrictions that an unprepared ERT context had on teachers' practice, ERT has also increased social inequalities among students (Martins, 2020), including students with special needs. As students had access to a wide variety of educational solutions, when they had a solution at all; social inequalities among students had also an impact on students' learning and performance. Subsequently, this impact on students' learning and performance had an impact on student's assessments. The ERT impact on assessments is particularly worrying when assessments are used with summative and grading purposes. When decisions about students' paths rely on assessments, as in grade transition or in the access to higher education, issues about the fairness of assessments arise. As the American National Academy of Education points out, an equitable assessment system is fair, accurate, and valid. In such a disruptive context, where fairness, reliability and validity of assessments are questioned, making use of assessments for accountability purposes became controversial (NAE, 2021; Nisbet & Shaw, 2022).

Although the core of Decree-Law No. 54/2018 and Decree-Law No. 55/2018 encourages formative assessment practices in the Portuguese education system, summative assessments also take a very important role on the system. Considering ERT limitations and impact on students' learning and assessments, the Portuguese government approved, in April 2020, a set of measures to establish an exceptional and temporary assessment scheme, regarding internal assessment, national assessment tests and basic and secondary education examinations, for the conclusion of the 2019/20 school year (Decree-Law No. 14-G/2020).

It should be once again noted that Decree-Law No. 14-G/2020 was written with clear concerns regarding summative assessments, grading and students' learning certification. To deal with ERT and lockdown problems, the Portuguese Ministry of Education authorized schools and teachers to adjust their classroom assessment methods, criteria, and instruments to students and to the circumstances. Moreover, all national assessment tests on grades 2, 5 and 8 were cancelled. Grade 9 examinations for basic education conclusion and grade 11 and 12 examinations for secondary education conclusion were also cancelled. Secondary education examinations were performed only as a condition to access higher education. Consequently, grading and learning certification, on all grades of compulsory education, only considered internal assessment. Therefore, this quick response to an evolving and changing context, and the displacement of summative large-scale assessments in favour of formative teacher-constructed assessments, can be defined as emergency assessment (Cooper et al., 2022).

Note as well that grading in each subject area should have taken as a reference the whole school year, including the time performed in ERT. To do so, schools had the autonomy to define distance learning methods that they considered most appropriate, paying attention to all available resources, assessment criteria, and the contexts in which students lived. To students of all grades, it was required to attend synchronous sessions, when these existed, and when students had technological instruments to access them. It was also required to participate with the proposed activities for the asynchronous sessions, with the duty of sending their homework to their teachers so that teachers could collect evidence for final summative assessment and grading (Decree-Law No. 14-G/2020).

In such a disruptive context as the ERT was for instruction, teachers' conceptions and practices about assessment were also confronted with a totally new paradigm. Panadero et al. (2022) showed that, in Spain, the sudden transition to ERT affected teachers' assessment practices as they were compelled to change their instruments, standards, and demands of students. Teachers became more flexible in their evaluation strategies, lowering their assessment criteria and standards; and amending grading procedures (Panadero et al., 2022).

In line to what has happened in Spain (Panadero et al., 2022), the study conducted by the CNE – *Conselho Nacional de Educação* [National Education Council] reveals that, in Portugal, the ERT raised on teachers doubts and questions about assessment methods, instruments, criteria, objects and trust. In this study, about 82% of teachers that participated in the survey indicated that assessment of learning was difficult or even very difficult to perform (CNE, 2021a).

Furthermore, Cooper et al. (2022) refers that, in Canada, as well as in Portugal, the pandemic has resulted in further responsibility being placed on teachers to leverage classroom assessment as the source for information about student growth and achievement. Such a shift has anchored assessments most closely to students' learning contexts and experiences; and it has forced teachers to become a crucial agent in assessment (Cooper et al., 2022).

Consequently, it is believed that part of the difficulties felt by teachers on their own adaptation to ERT are related to instruction and assessment, with an increased workload to adjust their instruction and assessment objects, methods, instruments, and criteria. Thus, we predicted that:

H2: Instruction and assessment (IA) mediate the relationship between teachers' adaptation (TA) and workloads (WL).

Figure 20. Conceptual model for Hypothesis 2



4.1.3. Changing Assessment Methodologies During ERT (The Moderator)

With the implementation of Decree-Law No. 14-G/2020, it was established a new legal framework that allowed teachers to adjust their assessment practices to ERT. According to CNE' study (2021a), 81% of Portuguese teachers that participated on their survey, have changed their assessment methodologies to adapt themselves to ERT. To adjust their instruction and assessment practices to remote teaching, teachers reshaped assessed objects, assessment criteria, assessment instruments, and assessment purposes.

In addition, Rodrigues et al. (2022) reveal that Portuguese teachers have also diversified their assessment instruments during ERT, and they have used less traditional written tests than before the ERT context. Consequently, it is expected that having changed assessment methodologies moderates the relationship between difficulties on teachers' adaptation and difficulties on instruction and assessment, in the sense that the relationship is more intense for teachers who did not change their assessment methodologies during ERT, leading to the following hypothesis:

H3: Having changed assessment methodologies (CAM) moderates the relationship between teachers' adaptation (TA) and instruction and assessment (IA), in the sense that the relationship is more intense when assessment methodologies have not been changed.

Figure 21. Conceptual model for Hypothesis 3



Cumulatively, the above-mentioned predictions suggest that the mediation role of instruction and assessment (IA) in the relationship between teachers' adaptation (TA) and workloads (WL) depends on the level of changed assessment methodologies (CAM), leading to this hypothesis:

H4: The indirect effect between teachers' adaptation (TA) and workloads (WL), via instruction and assessment (IA), is moderated by changed assessment methodologies (CAM).

Figure 22. Conceptual model for Hypothesis 4



4.1.4. Assessment in Special Education During ERT

According to Decree-Law No. 54/2018 for inclusive education, schools must ensure that all pupils have the right to participate in assessment processes and teachers should adjust their internal assessment processes to each student's needs.

For special education students' assessment, as the population of students is a very heterogeneous group, learning targets are often individualized. Still, individualized learning targets are linked to the general education curriculum and instruction. These individual learning goals should be well documented. Then, the general education teacher and the special education teacher collect formative assessment data to evaluate and monitor the student progress and, if necessary, to modify their learning goals (Xu, 2013). It should be noted that, although most of these adaptations have met the inclusion program needs, special education teachers often disregard the recommendation to grade their students on achievement (Brookhart, 2013). As teachers have their own conception and beliefs about assessment purposes and practice, these beliefs often superimpose themselves to recommendations and legislation regarding assessment and grading (Cipriano & Martins, 2021). Thus, on the one hand, grades should be reflective of students' achievement. On the other hand, most special education teachers place most emphasis on students' improvement than on students' achievement (McLeskey & Waldron, 2002).

During ERT, Womack & Monteiro (2022) have shown that special education services continued to be delivered. However, according to the students and caregivers' perspectives, the Covid-19 pandemic

negatively affected the educational performance of SEND students, namely on speech and language development, learning, and academic achievement (Pittas & Papanastasiou, 2023), and special education professionals reported lower levels of school connectedness and self-efficacy in their roles when compared to typical school year. Additionally, according to Schuck & Lambert (2020), special education' teachers described a tension around accountability, grading and attendance. They knew that students needed accountability and feedback but were unsure of how to go about providing it.

Therefore, as a complementary analyses, considering the above-mentioned information, we decided to run the model shown in Figure 22 on special education teachers only, to understand the relations between their own adaptation, perceived workloads, instruction and assessment, and changing assessment methodologies.

4.2. Method

As referred to in Section 3.2, mediation and moderation analyses were performed (Hayes, 2022) to test these formulated hypotheses using the database created by CNE to support the study *"Educação em tempo de Pandemia: Problemas, respostas e desafios das escolas"* [Education in Pandemic Times: Problems, Responses, and Challenges of Schools] (CNE, 2021a). From this database, different items were selected to construct composite variables for testing, i.e., to construct Teachers' Adaptation (TA), Workloads (WL), Instruction and Assessment (IA) and Changes on Assessment Methodologies (CAM) variables.

4.2.1. Measures

4.2.1.1. Teachers' Adaptation

As referred before within this chapter, teachers' adaptation to ERT implied the realization of several new tasks to be performed by students and teachers (E.g., CNE, 2021a; Flores & Gago, 2020; OECD 2020b, 2021). To measure this dimension of teachers' adaptation (TA) to ERT, a composite variable was constructed considering 3 items to the question *"As a coordinator, indicate the degree of difficulty that the following aspects had a*) *Adjust the number of tasks to students b*) *Adjust the complexity of tasks to students c*) *Plan weekly/fortnight tasks"*. All the 3 items were answered on a four-points scale ranging from 1 (very easy) to 4 (very difficult). TA composite variable was constructed with a very good reliability (WLA' α = 0.79, Kline, 2016), where higher scores indicate a higher level of difficulty on teachers' adaptation.

4.2.1.2. Workloads

Teachers also had an induced feeling of an increased workload during emergency remote teaching (CNE, 2021a; Klusmann et al, 2023) due to intensification and extensification of work efforts of homebased teleworking (Palumbo, 2020). To measure workloads (WL), a composite variable with an adequate reliability was constructed (α = 0.79, Kline, 2016), with higher scores indicating higher level of workload. WL was constructed considering eight items to the question *"Regarding the volume of work required by remote teaching, how do you classify the following tasks?"*. All the eight items were answered on a four-points scale ranging from 1 (less) to 4 (much more).

4.2.1.3. Instruction and Assessment

It is known that meaningful classroom assessment is strongly linked to the curriculum, instruction and learning processes (E.g., Black & Wiliam, 1998a, 2018; Fernandes, 2011; Lingard, 2007). To measure instruction and assessment (IA) considering these dimensions, a composite variable was constructed considering five items to the question "As a coordinator, indicate the degree of difficulty that the following aspects had a) Comply the syllabus, b) Diversify teaching methodologies, c) Promote autonomous learning, d) Provide feedback to tasks, and e) Assess learning". All the items were answered on a four-points scale ranging from 1 (very easy) to 4 (very difficult). As the IA Cronbach alpha was 0.68, the mean of inter-item correlation was also calculated (0.30) and reliability was assured (Clark and Watson, 1995). Higher scores of the new composite variables indicated a higher level of difficulty on Instruction and Assessment. It should be noted that, performing a factor analysis, all the five items remain on the same factor, with loadings ranging from .504 to .687. When variables within factors are highly intercorrelated, as in this case, factors are assumed to represent dimensions within the data, corresponding to concepts that cannot be adequately described by a single measure (Field, 2013; Hair et al. 2006). Therefore, as the five items remained in the same factor, it is assumed that the five items are measuring the same concept, i.e., instruction and assessment (IA). Thus, this revels that, even in a disruptive ERT context, a strong link between curriculum, instruction, assessment and learning processes remain.

4.2.1.4. Changes on Assessment Methodologies

To construct changes on assessment methodologies, it was considered answers to the question: *"While schools were closed, did you change assessment methodologies?"* Answers to this question (Yes or No) were recoded as a dummy variable – Changed Assessment Methodologies (CAM) – where (1) represents changed assessment methodologies and (0) has not changed.

4.2.2. Data Analysis Procedures

Descriptive statistics (mean and standard deviation) and correlations for study variables were calculated. The research hypotheses were tested on SPSS (Statistical Package for the Social Sciences – a software program for quantitative analysis of complex data) using the macro PROCESS created by Hayes (2022) to test mediation and moderation models.

The tests were performed on the whole sample and in special education sub-sample. For hypotheses 1 and 2 about mediation, it was used model 4, and for hypotheses 3 and 4 about moderated mediation it was used model 7 (Hayes, 2022). To assess the indirect effect and the conditional indirect effect, a bootstrap estimation was implemented, using 5 000 bootstrap samples, and confidence intervals at 95% were calculated. In addition, to interpret the moderated effect, the interaction effect was plotted at -1/+1 SD from the mean of the moderator changed assessment methodologies (CAM) (Aiken & West, 1991).

Due to age bias between the sample and population, age was controlled for in the mediation and moderate mediation models. As the results were similar, it was decided to report only results without age as a covariate.

4.3. Results

4.3.1. Preliminary Analysis

Descriptive statistics was used to analyse the model variables – Changed Assessment Methodologies (CAM), teachers' adaptation (TA), instruction and assessment (IA), and workloads (WL) (Table 16).

	Mean	SD	1	2	3	4
1. CAM ^a	0.83 ^b	-				
2. TA	2.45	0.53	.03	(.79)		
3. IA	2.77	0.46	.08*	.47*	(.68)	
4. WL	3.42	0.46	.10*	.22*	.24*	(.79)

Table 16. Descriptive statistics, reliabilities, and study variables correlations

Note. N = 3770. Cronbach's alpha is reported in parentheses

^a 0 = No change assessment methodologies, 1 = Change assessment methodologies.

^b Proportion of category 1 was reported

* *p* < .001

Source: CNE teachers' survey, 2020; Own calculations

4.3.2. Hypotheses Testing Considering all Curricular Areas

Hypothesis 1 regarding the positive relationship between teacher's adaptation (TA) and workloads (WL) was supported (B = 0.19, t = 13.90, p < .001). Teachers' adaptation (TA) had a significant effect on workloads (WL), mediated by instruction and assessment (IA), and the 95% confidence interval for the indirect effect did not include 0 (B = 0.08, 95% CI = 0.06, 0.09, Table 17). This supports H2. Thus, instruction and assessment (IA) explains the process by which teachers' adaptation affects workloads.

					R ²	
Model 1: mediator variable in the model	Outcome: Instruction and Assessment (IA)					
	Coeff.	SE	t	p	.22	
Teachers' adaptation (TA)	0.40	0.01	32.75	< .001		
Model 2: outcome variable in the model	Outcome: Workload (WL)					
	Coeff.	SE	t	p	.07	
Teachers' adaptation (TA)	0.12	0.02	7.65	< .001		
Instruction and assessment (IA)	0.18	0.02	10.19	< .001		
	Indirect effect (via Instruction and assessment)					
	Bootstrapping 95% CI					
	Effect	SE	LL	UL		
Indirect effect of Teachers adaptation on workloads via Instruction and Assessment	0.08	0.01	0.06	0.09		

Table 17. Results for mediation considering all curricular areas

Notes. N = 3786. Unstandardized regression coefficients are reported. 5000 bootstrap samples. LL – Lower Limit; UL – Upper Limit; Cl – Confidence Interval.

Source: CNE teachers' survey, 2020; Own calculations

Hypothesis 3 predicted that having changed assessment methodologies (CAM) moderates the relationship between teachers' adaptation (TA) and instruction and assessment (IA), and the relationship is more intense when assessment methodologies have not been changed. Testing results show that the interaction effect between teachers' adaptation (TA) and having changed assessment methodologies (CAM) is negative and significant (B = -0.09, t = -2.90, p = .004, Table 18).

Table 18. Results for moderated mediation considering all curricular areas

					R ²	
Model 1: mediator variable in the model	Outcome: Instruction and Assessment (IA)					
	Coeff.	SE	t	p	.23	
Teachers' adaptation (TA)	0.48	0.03	16.50	< .001		
Changed assessment methodologies (CAM)	0.08	0.02	4.33	< .001		
TA x CAM	-0.09	0.03	-2.90	.004		
Model 2: outcome variable in the model	Outcome: Worklo	Outcome: Workload (WL)				
	Coeff.	SE	t	p	.07	
Teachers' adaptation (TA)	0.12	0.02	7.65	< .001		
Instruction and assessment (IA)	0.18	0.02	10.19	< .001		
	Conditional indire	ect effect (via In	struction and	assessment)		
			Bootstra	ap 95% Cl		
	Effect	SE	LL	UL		
No change assessment methodologies (0)	0.09	0.01	0.07	0.11		
Change assessment methodologies (1)	0.07	0.01	0.06	0.08		
Index of moderated mediation	-0.02	0.01	-0.03	-0.01		

Notes. N = 3771. Unstandardized regression coefficients are reported. 5000 bootstrap samples. LL – Lower Limit; UL – Upper Limit; Cl – Confidence Interval.

Source: CNE teachers' survey, 2020; Own calculations

Additionally, the relationship between teachers' adaptation (TA) and instruction and assessment (IA) is positive and significant when the assessment methodologies have not changed (B = 0.48, t = 16.50, p < .001, Figure 23). This relationship is also positive and significant but weaker for changed assessment methodologies (B = 0.38, t = 28.12, p <.001, Figure 23).

Figure 23. Moderation effect of changed assessment methodologies in the relation between teachers' adaptation and instruction and assessment



Source: CNE teachers' survey, 2020. Own calculations

The results also showed that changed assessment methodologies (CAM) significantly moderated the indirect effect of teachers' adaptation (TA) on workloads (WL), through instruction and assessment (IA) (bootstrap estimate = -0.02, Boot CI = -0.03, -0.01). For those who have not changed their assessment methodologies, teachers' adaptation (TA) had a significant indirect effect on workloads (WL), through instruction and assessment (IA) (B = 0.09, Boots 95% CI (0.07, 0.11)). For those who have changed their assessment methodologies, the indirect effect was weaker (IA) (B = 0.07, Boots 95% CI (0.06, 0.08)).
4.3.3. Complementary Analysis: Special Education Area

The same hypotheses were tested in the sub-sample for the special education curricular area. Hypothesis 1, concerning the positive relationship between teacher's adaptation (TA) and workloads (WL), was supported (B = 0.19, t = 3.94, p = .001). Teachers' adaptation (TA) had a significant effect on workloads (WL), mediated by instruction and assessment (IA), and the confidence interval for the indirect effect did not include 0 (B = 0.10, 95% CI = 0.04, 0.17, Table 19), supporting hypothesis 2. Thus, instruction and assessment (IA) explains the process by which teachers' adaptation affects workloads, also in special education curricular area.

					R ²	
Model 1: mediator variable in the model	Outcome: Instruction and Assessment (IA)					
	Coeff.	SE	t	p	.27	
Teachers' adaptation (TA)	0.41	0.04	9.75	< .001		
Model 2: outcome variable in the model	Outcome: Workload (WL)					
	Coeff.	SE	t	p	.09	
Teachers' adaptation (TA)	0.08	0.06	1.34	.182		
Instruction and assessment (IA)	0.25	0.08	3.25	.001		
	Indirect effect (via Instruction and assessment)					
Boots					I	
	Effect	SE	LL	UL		
Indirect effect of Teachers adaptation on workloads via Instruction and Assessment	0.10	0.03	0.04	0.17		

Table 19. Results for mediation for special education curricular area

Notes. N = 255. Unstandardized regression coefficients are reported. 5000 bootstrap samples. LL – Lower Limit; UL – Upper

Limit; CI – Confidence Interval.

Source: CNE teachers' survey, 2020. Own calculations

Since having changed assessment methodologies (CAM) was not a significant moderator (Table 20), hypotheses 3 and 4 were not supported.

Table 20. Results for moderated mediation for special education area

					R ²	
Model 1: mediator variable in the model	Outcome: Instr	Outcome: Instruction and Assessment (IA)				
	Coeff.	SE	t	p	.27	
Teachers' adaptation (TA)	0.42	0.04	5.71	< .001		
Changed assessment methodologies (CAM)	0.06	0.05	1.24	.216		
TA x CAM	-0.03	0.09	-0.28	.780		
Model 2: outcome variable in the model	Outcome: Worl	Outcome: Workload (WL)				
	Coeff.	SE	t	p	.09	
Teachers' adaptation (TA)	0.09	0.06	1.40	.165		
Instruction and assessment (IA)	0.27	0.08	3.44	.001		
	Conditional indirect effect (via Instruction and assessment)					
	Bootstrap 95% Cl					
	Effect	SE	LL	UL		
No change assessment methodologies (0)	0.11	0.04	0.05	0.19		
Change assessment methodologies (1)	0.11	0.03	0.05	0.17		
Index of moderated mediation	-0.01	0.03	-0.06	0.05		

Notes. N = 253. Unstandardized regression coefficients are reported. 5000 bootstrap samples. LL – Lower Limit; UL – Upper Limit; Cl – Confidence Interval.

Source: CNE teachers' survey, 2020. Own calculations

4.4. Discussion

The research carried out in this chapter showed that there was a relationship between teachers' adaptation to emergency remote teaching and perceived workloads. This finding is in line with other researches that refers that working from home increased workload perception (CNE, 2021a; Klusmann et al, 2023; Palumbo, 2020).

In addition, the results showed that instruction and assessment mediated the relationship between teachers' adaptation and perceived workloads. As Cooper et al. (2022) referred, the pandemic has resulted in additional responsibility being placed on teachers' assessments as the source for information about student development and accomplishment. Such a shift has forced teachers to become a crucial agent in assessment. Also, it should be noted that, when performing classroom assessment, assessment purposes and data collection procedures should be directly related to the type of interaction that a teacher establishes with their students (Fernandes, 2013; Stufflebeam & Coryn, 2014). With the sudden transition to ERT, in the presence of a different pedagogical relationship, it was expected that teachers' adaptation to ERT would imply heavier workloads to adjust their instruction and assessment methods to this new context. Therefore, as predicted, teachers own adaptation and perceiver workloads during ERT are mediated by instruction and assessment, both on the sample and special education sub-sample.

When all curricular departments were considered, changing assessment methodologies, legally proposed by Decree-Law No. 14-G/2020, moderated the relation between teachers own adaptation to ERT with instruction and assessment. Thus, the mediation role of instruction and assessment in the relationship between teachers' adaptation and workloads depended on the level of changed assessment methodologies. This means that this legal framework was an important policy to allow teachers a better fitting to this new exceptional context.

It is important to note that, often, public policies guidelines have different interpretations and implementations, in different contexts, which do not always translate into a direct, coherent, and obvious practice (Ball, 2008). This becomes evident on this sub-study when, considering special education teachers only, it was showed that changing assessment methodologies proposed by Decree-Law No. 14-G/2020 did not moderated the relationship between teachers own adaptation to ERT with instruction and assessment, as it has happened when all curricular areas were considered. This means that this legal framework, regarding changes on assessment of learning during the pandemic, had no significant effect on instruction, assessment and perceived workloads for special education teachers; and difficulties felt by special education teachers on their own adaptation went beyond changing assessment methodologies.

These findings are in line with other studies that shows that special education teachers had lower levels of school connectedness and self-efficacy, with a tension around accountability, grading and attendance (Schuck & Lambert, 2020; Womack & Monteiro, 2022). We have to consider that previous adjustments on assessment to each student individual contexts (as provided for in Decree-Law No. 54/2018; and underlined in Xu, 2013) and the emphasis that special education teachers tend to put into improvement when assessing (Brookhart, 2013; McLeskey & Waldron, 2002) might justify why the relation between teachers' adaptation and instruction and assessment is not moderated by changing assessment methodologies. Nevertheless, further research on this topic should be developed to better understand these relationships during the ERT.

When it comes to the implications of these findings, firstly, when considering all curricular areas, the investigation carried out in this chapter showed that changing assessment methodologies during the first ERT played a vital role in moderating the relationship between teachers' adaptation, instruction and assessment, and perceived workloads. Due to a shift in the educational context, a legal

115

framework allowing changes in assessment methodologies was an important policy to allow teachers a better fitting to this new exceptional context. While the influence of contexts on students' learning has already been considered in alternative assessment models for special education students (Xu, 2013), this study contributes to a broader and essential reflection on the need to consistently consider learning contexts in classroom assessments, not only in special education but across all forms of education (Brown, 2022; Cooper et al., 2022; NAE, 2020). Therefore, the Covid-19 pandemic period highlights the importance of understanding and documenting learning processes and contexts, and the need to account for them in the design and interpretation of assessments, both in classroom and largescale assessments (NAE, 2020).

Secondly, this research showed that changes in assessment methodologies proposed by Decree-Law No. 14-G/2020 during the first lockdown had varying impacts on moderating the relationship between teachers' adaptation to instruction and assessment, as well as their perceived workloads. This reinforces the need for policymakers to consider that policies can be interpreted and implemented differently in various contexts (Ball, 2008), leading to varying outcomes. Consequently, when creating new policies, it is crucial for policymakers to account for the diversity of educational contexts and recognize that the same policy can result in different implementations and impacts.

Lastly, it is also known that changing assessment practices during the Covid-19 pandemic is associated to heavier perceived workloads (CNE, 2021a). In addition, as recalled by Panadero et al (2022), it has been long acknowledged that we need to improve our teachers' assessment literacy. We hope that this study might bring new insight about Portuguese teachers' assessment practices during the pandemic and its relationship with perceived workloads; as work efforts and workloads to perform classroom assessment should also be considered in future professional development programs to avoid consequences in teachers' health conditions (Lizana & Vega-Fernadez, 2021).

4.5. Chapter Final Consideration

The social disruption caused by the covid-19 pandemic on education systems has stimulated a healthy discussion about their ability to adapt to different contexts and their capacity to recognize problems and create solutions in situations of adversity (OECD, 2020a). With the sub-study within this chapter, it becomes clear that Portuguese teachers were able to adapt their classroom assessment practices, and that this process had implications on their perceived workloads.

It also becomes clear that difficulties felt by special education teachers on their own adaptation to ERT went beyond issues related to changing assessment methodologies. Therefore, it is important to develop further investigation on this topic, to be able to create and provide measures and mechanisms that truly ensure inclusion and access to education to all students, with their full right to participate in assessment processes, even in disruptive emergency contexts as the covid-19 lockdowns were. This page intentionally left blank

CHAPTER 5

The Washback Effect of External Assessments and the Covid-19 Pandemic: Perspectives of Head Teachers in Portugal

Chapter Summary:

External assessment of learning has been the subject of very different perspectives, conceptions, and assumptions. Considering the pressure that the existence of external assessments can generate over the functioning of education systems, some recognize benefits, and others point out negative and perverse effects of this pressure. In 2020, in most countries, the Covid-19 lockdowns involved the cancelation of large-scale external assessments in favour of teacher-built assessments. In 2023, a full reintroduction of external assessment of learning was discussed in Portugal, and the role of external assessments in the Portuguese education system was a hot topic of discussion in educational communities. In this post-pandemic context, where the pressure of external assessments has been temporarily relieved, for the writing of this fifth chapter, it was sought to know the point of view of Portuguese head teachers about the impact of this temporary external assessment suspension in schools. Moreover, it was sought to know, from the head teachers' point of view, which should be the external assessment model to be implemented for the future. To do this, a series of 32 interviews were conducted with head teachers from all over the country. Results show that the pandemic has challenged the purposes and format of external assessments, highlighting different positions of value regarding its reintroduction in the Portuguese context.

5.1. External Assessments

External assessment of learning usually involves the application of a common test, based on prescribed syllabi in curriculum, with emphasis on written tasks, serving several functions, and it is set and/or controlled by an agency external to the schools from which candidates come from, at a state or national level (Kellagham & Madaus, 2003). External assessment tests, as examinations, have been employed for a variety of reasons and purposes, such as a form for policymakers to influence instruction and learning in schools, and a form of accountability (Madaus & Russell, 2010). With a strong ideological orientation due to certain representations of school, over the years, along with

school systems development in the world, educational stakeholders have had different perspectives, conceptions, and assumptions regarding external assessment tests (Machado et al, 2022). Hence, in most countries in the world, external assessment tests continue to play a significant role in education systems. Although their real pedagogical value can be questioned, policymakers continue to insist on their use for reasons that are often associated with the idea that they are a credible measure for quality of teaching, quality of learning and, in general, quality of education (Fernandes, 2019).

5.1.1. Consequences of External Assessment Tests in Education Systems

Quite often, literature refers to external assessment tests considering their associated stakes, referring to them as low-stakes assessments or high-stakes assessments (cf. Rozenwajn & Dumay, 2014; Stobart & Eggen, 2012). On the one hand, low-stakes assessments aim to influence pedagogical practices towards the improvement of teaching and learning quality, providing feedback to stakeholders. Results from low-stakes assessments do not have formal consequences for schools, teachers, or pupils. Here, the reflexive logic promoted by the external assessment tests' results should be sufficient to produce a process of improvement. On the other hand, with high-stakes assessments, results are associated with quite important consequences for schools, teachers, or pupils. In this case, it may involve a risk of dismissal of head teachers in the event of recurring unsatisfactory results; and it may involve grade repetition or the refusal to access higher education for students depending on their final grades. Here, the driving force behind the pedagogical change is the motivation of individuals to avoid sanctions or to obtain rewards (cf. Rozenwajn & Dumay, 2014; Stobart & Eggen, 2012).

According to Alderson & Wall (1993), tests are held to be powerful determiners of what happens in classrooms; and the decisions about the use and format of external assessment tests are indelibly linked to the political and ideological choices of those who have the responsibility to put them into practice (Fernandes, 2019). High-stakes testing, when used as a policy tool, can promote a variety of values that include utilitarianism, economic competitiveness, technological optimism, objectivity, bureaucratic control, accountability, administrative convenience, numerical precision, efficiency, standardisation, and conformity (Madaus & Russell, 2010). However, it seems that high-stakes testing can also provoke negative reactions from educational actors, leading to the development of perverse effects that negatively affect learning processes. Negative effects of high-stakes testing include narrowing the curriculum, decreasing attention on non-tested subjects, narrow test preparation, corruption of test results, cheating, retaining students in a grade, increased dropout rates, and increasing student stress and anxiety (Kellagham & Madaus, 2003; Madaus & Russell, 2010; Mons, 2009; Mustin, 2017). In educational assessment, washback refers to the phenomenon when testing (often large scale and high-stakes), specifically the uses of test scores and the decisions made based on those scores, influence those stakeholders associated with such testing and potentially their teaching and learning (Tsagari & Cheng, 2017). It is important to note that it is not the test itself that causes the washback phenomenon and negative disorders. Negative impacts do not derive from any source of test invalidity such as construct under-representation or construct-irrelevant variance (Messick, 1996). Test impacts relate more directly to use than to the interpretation of scores (Gipps, 2011); and it is the stakes associated with test scores that drive teachers, pupils, and other stakeholders into behaviour that results in the many unintended outcomes (Madaus & Russell, 2010). Moreover, Fernandes (2019) highlights that the greater the consequences are given to test scores, the greater the tendency to narrow the curriculum, deepening harder these negative effects.

5.1.2. Consequences of External Assessment Tests in Portugal

Fernandes (2019) states that *Provas de Aferição* (PA) can be considered very low-stakes assessments; *Provas Finais de Ciclo* (PFC) can be considered low-stakes assessments; and *Exames Nacionais* (EN), in secondary education, can be considered medium-stakes assessments, as they can have a significant weight to access higher education. Formally, medium or high consequences exist only for students by the end of secondary education, who may have to repeat a school year, who may be unable to access higher education, or who may be unable to choose the course they want to do in higher education. For all the other educational agents, the results of PA, PFC, and EN do not imply formal consequences. In extreme cases, inspection services can start disciplinary proceedings; but even that is quite rare.

Regardless of the low-, medium-, or high-stakes label that might be given to external assessment tests on policies, it should be noted that a same test can have different stakes depending on the stakeholders' corporative and/or personal goals. Therefore, the same test results can have different consequences for educational actors at the same level. For example, for students, EN have different stakes depending on students' willingness to enrol (or not) in higher education, and the course they want to attend. At a school administration level, Torres et al. (2019) highlighted that the adoption of mechanisms to promote academic distinction of the best students based on students' results, and the centrality of EN at the end of secondary education, have redirected everyday teaching in schools towards the achievement of academic performance goals by students (Torres et al., 2019), with different stakes associated considering head teachers corporative and/or personal goals.

Moreover, as consequences of external assessment tests in Portugal, Machado et al. (2022) reveal that the aspects that raise most criticism about EN among Portuguese head teachers are their relationship with access to higher education, their reliability and "fairness", and their effects on the school rankings creation, constructed through the results obtained by the students. Nevertheless, while the rationale of PFC and EN is the object of consensus among head teachers, namely the necessity of their existence, their certifying function, and their function as an element for decision-making concerning students' approval and selection; the rationale of PA, on the contrary, is not the object of appropriation and operationalisation. Some head teachers even consider PA a waste of time. Accordingly, Cipriano & Martins (2021) also reported that at a teaching level, teachers in Portugal tend to reject the existence of PA, and they tend to approve the existence of PFC and EN, especially EN in secondary education. Not only do they agree more with the existence of PFC and EN, but they also consider that these are more important than PA. Mathematics and Science teachers are the ones who value the existence of PFC and EN the most.

Considering the effects of external assessment tests and its implications to the education system functioning, before the pandemic, the Portuguese National Council of Education (CNE) had already written a recommendation saying that the weight that EN have to access higher education, create a pressure with negative effects on secondary education objectives (Lourtie, 2020), considering the "Students profile by the end of compulsory education" (see PASEO, 2017). But suddenly, the Covid-19 pandemic came, resulting in the cancellation of all PA and PFC. EN in secondary education were performed only to access higher education (see Decree-Law No. 14-G/2020).

5.1.3. Research Opportunities and Questions

With the research carried out to write this chapter, it was sought to get to know the perspectives of Portuguese head teachers regarding a) this minimum model of external assessment implemented during the Covid-19 pandemic and its impacts on schools functioning, b) the impacts of external assessments in internal practices, and c) what should be the external assessment model to come after the Covid-19 pandemic. Consequently, two sub-research questions were formulated:

RQ 2.1. Was the cancelation of external assessment of learning during the Covid-19 pandemic a good decision? Why?

RQ 2.2. Should external assessment of learning be reintroduced for the conclusion of basic and secondary education? Why?

5.2. Method

As referred in section 3.3, to be able to capture local educational context, as well as the larger social, political, and economic factors governing teaching and learning in relation to a test/examination or a testing system proposed by Tsagari & Cheng (2017), a qualitative research approach was designed, with 32 semi-structured interviews, conducted by videoconference calls with head teachers from different types of school and regions in Portugal. The interview script had 16 open-ended questions about the impact of the Covid-19 pandemic on assessment practices, intending to elicit views and opinions from the participants in this topic (Creswell & Creswell, 2023).

5.3. Results

5.3.1. Preliminary Analysis

A preliminary analysis was conducted considering the three documents' variables: i) School cluster type, ii) Administrative and governance characteristics, and iii) Region by Nomenclature of Territorial Units. Although content variations are very similar regardless of these three variables, it was possible to observe on head teachers' speeches from schools with basic education only, a prevalence of segments regarding the need to restructure PA, when compared with other schools' head teachers. In addition, it was possible to observe, for head teachers from TEIP schools, a prevalence of segments regarding how external assessment devalues students' paths in school and devalues the work that is done by teachers, when compared to other schools' head teachers speeches. No significant content variations were found when considering NUTS II.

5.3.2. Inequalities in ERT Provision

At the outset for this content analysis, it must be noted that the context in which schools were functioning underwent a major shift in a couple of weeks and, as a result, head teacher reported that they faced many inequalities in providing Emergency Remote Teaching to students:

> In our school, we have students in both scientific-humanistic courses [general secondary education] and vocational courses. ERT in the vocational courses was more complex to implement because these students had to undergo training in a work environment, such as internships. However, with establishments closed, we had to be creative; otherwise, they would not complete their educational path. These are students who either finish school that year, or we would lose them

because they would not return the following year. (...) These students have different characteristics, and they do not return to school to repeat a subject or a module; they do not come back (HT 28).

We are a school associated with prison establishments; we have about 300 students there in the two prison facilities. There, at prisons, it was much more complicated and much slower to implement ERT (HT 30).

The good students survive to this and much more (...). Then we have the students who have problems and difficulties. These are the biggest victims (HT 22).

5.3.3. Cancelation of External Assessment During the Covid-19 Pandemic

Regarding RQ 2.1. – Was the cancelation of external assessment of learning during the Covid-19 pandemic a good decision? Why? – there is almost a consensus among head teachers that cancelling PA, PFC, and EN for basic and secondary education conclusion during the pandemic was a good and inevitable decision. Educational inequalities, public health, physical space in schools, human resources for surveillance, and students' emotional well-being conditions were not met to perform PA, PFC, and EN with all students. So, the cancelation was inevitable and, therefore, it was a good decision:

I would say that the difficulties were so obvious that it was impossible to be otherwise (HT 21).

At the time, it was avoided, as much as possible, gatherings and the presence of people in physical spaces where the virus could potentially spread most easily. Therefore, from that point of view, I think it was a good decision (HT 22).

Moreover, having EN just to access higher education has worked, and it was appropriate for the circumstances. Eventually, even EN could have been completely cancelled, but that would have created a major problem for higher education access:

External assessment goal is to have a standardized assessment applied to all. But with the pandemic, the responses that were given to students by schools were not the same for everyone. It did not make sense to have a standardized test for the conclusion of basic and secondary education (HT 23). Eventually, EN could have not been performed at all (HT 1).

In addition, some head teachers consider that the cancelation of PFC and EN for basic and secondary education conclusion, in the following years, had positive consequences on schools:

The pandemic years have freed students from the strain of final summative assessment. I am talking mainly about secondary education. The tension that secondary education students have regarding their final grade mark (HT 5).

I think it made assessment lighter. It does not mean it is less demanding, okay? But it has made assessment a lighter process, a less unhealthy process (HT 32).

Nevertheless, some head teachers consider that there is lost information that should have been gathered sooner through external assessments, namely PA, and that would have been important:

I think external assessments could have been done because, maybe, we would have realized sooner what each student has lost in relation to learning. We did not have to wait a year and a half, with all these diagnostic external tests that we have been doing, which made us see, clearly, some deficiencies that were there. We would have realized earlier that we were not able to reach many students, and that there were things that students lost. Because they actually lost things: they lost parts of their school paths and today we are still paying for it (HT 18).

5.3.4. External Assessment Reintroduction

Regarding the reintroduction of external assessment in the post-pandemic period – (RQ 2.2. Should external assessment of learning be reintroduced for the conclusion of basic and secondary education? Why?) – the previous consensus is lost. Selecting the most developed categories in the content analysis (categories with $n \ge 14$ segments) to have a significant dimension to demonstrate variation with well-established relationships among the categories (Bryman, 2012: 241) and performing a relation of codes mapping by the occurrence of codes in the same document in MAXQDA, it was possible to identify three head teachers' profiles, shown in Figure 24.

According to Rädiker & Kuckartz (2020, pp. 94-96), the Code Map displays the similarities between categories in a two-dimensional visualization. It is based on the tabular representation in the Code Relations Browser. The more co-occurrences two codes have, i.e., the more similarly they are used in the data, the closer they are placed together on the Code Map. The positions on the map are calculated using classical multidimensional scaling, a method of multivariate statistics. Additionally, hierarchical cluster analysis with unweighted average linkage is performed to identify code groups.





Notes. PA – Provas de aferição; PFC – Provas finais de ciclo; EN – Exames nacionais; EA – External Assessment; PASEO – [Students' Profile by the End of Compulsory School]. Source: Primary data, 2023

Figure 24 displays, therefore, the similarities (defined here as co-occurrence within the same segment) among the categories regarding the reintroduction of external assessment. The sizes of the symbols (circles) in the Code Map represent code frequencies, and the colours of the codes correspond

to the calculated cluster assignments on the map. For example, "EN + Tension" was coded most frequently, followed by "External Assessment, Yes: to gauge" as they have the larger circles.

Note that as the more similarly categories are used in the data, the closer they are placed together on the Code Map. For example "EN, Yes: for certification," "EN, Yes: as a guarantee of learning quality," "EN: fairest," "EA with low stakes, but with stakes," "EN, Yes: for accountability," and "EN: Yes" are close to each other. In contrast, categories such as "PFC and EN, No" and "Higher Education should create its own access model" are positioned far from these. Thus, it was formed three clusters: the first centred around "Accountability", the second comprising categories related to "Improvement," and the third focused on the need to "Restructure" External Assessments.

Below, on the left side of Figure 24, in dark blue, it is represented head teachers who are favourable to the reintroduction of PFC and EN, and they tend to reject the existence of PA because they have no stakes for students as, for them, external assessments are important for accountability purposes. They consider that PFC and EN are important to preserve high standards in schools and they are a mean to guarantee the quality of learning. Moreover, PFC and EN are important for students' and schools' accountability, and they are the fairest instrument for the conclusion of basic education, the conclusion of secondary education and access to higher education, as they put all students at the same level at a national level. Regarding PA, these head teachers consider that they should have stakes associated for students. Otherwise, students and teachers do not care about these external assessment tests and, therefore, their results have low reliability:

The 9th grade PFC, in Portuguese and Mathematics subject areas, are fundamental. In the 9th grade, it is very difficult to motivate students to study, to get them involved, and PFC directs students to prepare themselves and to engage with school. Moreover, PFC have an interesting weight for basic education conclusion (...) and they allow students to be aware of their own knowledge, their skills, and their quality. PFC, by the end of 9th grade, for 14-year old students, are important as they give seriousness to that school grade, which no longer had with the pandemic model and PFC cancelation. In secondary education, with EN, the same happens (HT 16).

I think that PA should have some weight for classification purposes – it does not have to be a very big weight, but having some weight would give some responsibility to those who perform them – I think that would be important (HT 15).

With EN for the secondary education conclusion, students know that part of their final grade will depend on that EN. Therefore, they commit and engage to the test, and we have greater reliability in these results. And that is also a very

important issue: the reliability of EN's results to draw conclusions regarding schools and the education system. In my opinion, we are not able to draw the same conclusions, with the same reliability, with PA's results (HT 28).

In green, at the top right of Figure 24, it is represented head teachers who are favourable to PA reintroduction and against PFC and EN reintroduction as, for them, the sole purpose of external assessment should be improvement. They consider that PA can promote change and can give relevant and important information to improve students' learning, schools' practices, and the system's performance. They also consider that PFC and EN give primacy to certain disciplines, such as Portuguese and Mathematics, increasing parents' pressure for their children to perform well in PFC and EN:

For a long time, we have lived in a school system in which only Portuguese and Mathematics were important, as a consequence of PFC in these disciplines at the 4th, 6th, and 9th grade. When 4th and 6th grade PFC were banned, and PA were introduced in disciplines such as Music, Arts, and Sports in the 2nd grade, what we have noticed is that teachers, nowadays, are concerned with the full curricular matrix. Previously, the curricular matrix of 1st cycle was something like this: 7 hours of Portuguese, 7 hours of Mathematics, and then a few hours for expressions and complementary offers. And before the introduction of PA in the 1st cycle, as there was no type of external measurement or assessment for Music, Arts, and Sports, they were practically non-existent at schools. If you take a look at 1st cycle school buildings, most of them do not even have a sports hall. None of them has great conditions for Sports practice. Municipalities, themselves, did not value this area of knowledge very much. With the PA introduction in disciplines such as Music, Arts, and Sports, along with other educational policies but essentially with PA introduction, teachers and society began to value these disciplines (HT 24).

For these head teachers, PFC and EN do not allow to manage the curriculum in schools and classrooms, and it promotes a *"testing culture"* instead of the diversification of assessment instruments, widening the social inequalities among students. In addition, these head teachers consider that it would be important to change the way access to higher education is performed, and higher education institutions should create their own access model, ending PFC and EN in the Portuguese school system:

It causes me some discomfort that we are always under stress because of EN and the higher education access. We should be concerned about the quality of basic and secondary education, and students' success by the end of the 12th grade. (...) We should not be concerned about higher education access. I wish universities and polytechnical schools would have another perspective about this, but... (HT 20).

Nowadays, secondary education is mostly focused on EN performance to access higher education. It turns out to be, almost, a pre-stage of higher education (...). Since secondary education is compulsory, as school is compulsory up to the age of 18 years old, I think that EN should not be compulsory. It is a principle. I think there should be another mechanism to measure pupils' learning and the quality of learning, like PA, in addition to the internal mechanisms that we have. But without this examination weight to finish secondary education (HT 27).

Below, on the right side of Figure 24, in light blue, it is identified head teachers who consider that PA needs improvements, and PFC and EN should not have been put back in place for basic and secondary education conclusion. These head teachers recognize that it is important the existence of external assessments, and their existence should be mainly for improvement and gauging purposes. Although PA have several problems from different orders, they consider that PA must keep existing, and they have to be restructured:

> It is important to review the PA model. Perhaps, the Ministry of Education has already recognized that, and has realized that socio-economic contexts are also fundamental to be considered. PA, as they stand, are not enough (HT 10).

> For the 2nd, 5th, and 8th grade' PA, I do not see the need to perform them with all students. I think it could be done by sampling, and from there we could draw conclusions about what is happening in the education system (HT 21).

These head teachers also consider that PFC and EN should not be reintroduced for the conclusion of basic and secondary education because they negatively influence internal pedagogical and assessment practices, reflected in what is taught in schools, on internal assessment practices and internal tests' structure. These negative effects are stronger in secondary education with EN than in the 9th grade with PFC, as EN have higher stakes than PFC. Moreover, EN also negatively influence the PASEO objectives, increasing stress for all the educational community.

Hence, EN promote a school ranking based only on EN results, which is unfair to students and schools' accountability. Therefore, they consider that the pandemic model could be kept, with EN being performed just to access higher education, as it was a less harmful model:

PFC in the 9th grade, teachers know it may jeopardise, in rare cases, a grade transition situation, and they know that PFC will have little consequences for the student. Therefore, external tests' effects, in the 9th grade, do not get as noticed as in secondary education. In secondary education, we teachers feel that our students can jeopardise 2 or 3 years of their work and their future lives in those 90 minutes that last an EN (HT 4).

[Within a discipline with EN in secondary education], no teacher will dare to perform only one written test during a semester and use other internal assessment instruments to assess their students. On the following day, I would have emails dropping in from the parents – Why only one test? Will my child be under-assessed? – The test use as a privileged instrument to assess students is also supported by parents (...) and the influence of EN on internal assessment practices will continue to exist as long as EN tests are used to access higher education (HT 30).

EN are unfair to schools because of school rankings. It is a perfectly primitive thing. Although the Ministry of Education does not promote these rankings, letting them to be produced by the Media, and letting everyone to do whatever they want with it, it is almost as criminal as making them available (HT 2).

5.4. Discussion

As key ideas from these 32 interviews, along with previously identified profiles in the Results section, it is important to highlight that all head teachers considered that it is important the existence of external assessment. Hence, it was possible to note that head teachers' conceptions regarding external assessment have an intrinsic and contrasting association between PA and EN, as showed by Machado et al. (2022). Moreover, this association is relevant for the interpretation of head teachers' perceptions, since EN are generally the standard through which PA are assessed (Machado et al., 2022).

Most head teachers consider that PA have none or very little negative effects on stakeholders' practices, and secondary education EN create more dysfunctions than basic education PFC. Regarding EN and PFC, head teachers reported that it is not the test weight that creates dysfunctions (they have

the same weight of 30% in basic and secondary education), but it is the additional use that EN have in secondary education to access higher education that creates the dysfunctions. Moreover, most of these dysfunctions get to happen within a discipline, across subject areas, and across grades. This is in line with previous research referring that it is the stakes associated with test scores that drive teachers, pupils, and other stakeholders into behaviour that results in the many unintended outcomes (Gipps, 2011; Madaus & Russell, 2010); and that the greater the consequences are given to test scores, the greater are the dysfunctions (Fernandes, 2019).

Most head teachers also reported that the additional use of EN to access higher education is actually a high-stake use and, consequently, EN in secondary education become the driving force behind educational action instead of PASEO, the students' profile by the end of compulsory school that should be the beacon of all education system. This fact is in line with the recommendation made by the CNE (Lourtie, 2020), referring that EN create a pressure to prepare students as candidates for higher education, with negative effects on the pursuit of secondary education objectives given the PASEO.

For many head teachers, EN centrality to access higher education and consequent dysfunctions in compulsory education functioning is a problem that should be solved by higher education institutions (selection and access). In addition, this is a problem that is highly unfair for students who do not want to go to higher education. Therefore, when it comes to understand who gets to be most harmed with these negative effects, they say that EN get to be more harmful to low-achievement students who just want to finish secondary education and cannot because they fail at the EN. Eventually, they will be unable to finish secondary education, or they will finish secondary education later, in recurrent education system (2nd chance education for adults), attending school in the evening (post-labour time), already without EN, and this is more expensive for everyone. Consequently, several head teachers consider that the pandemic model of external assessment could remain, as it was less harmful, especially for low-achievement students:

A fact: The inexistence of EN to conclude secondary education has solved many problems for low-achievement students that were here unable to finish secondary education. They would manage to be internally approved with 10 or 11 points out of 20, but then they were condemned to grade repetition at EN and could not complete secondary education because of one exam. That was troubling. The pandemic model of external assessment has solved that situation (HT 12). Some head teachers also reported what they consider to be a great policy contradiction: Although it was approved, in recent years, important policies to change assessment cultures and practices towards learning improvement – such as: a) the PASEO (2017), a humanistic basis profile for the compulsory school conclusion; b) the Decree-Law No. 55/2018 (from 2018), a legal framework that puts formative assessment and curricular flexibility as the core of the Portuguese compulsory school system; c) the Decree-Law No. 54/2018 (from 2018), another legal framework seeking to promote inclusion of all students in the education system; and d) the MAIA project (since 2019), a teacher training program to improve teachers assessment literacy and practices (DGE, 2023); – it is already known that PFC and EN will be reintroduced in a post-pandemic context, in 2024, for the conclusion of basic and secondary education. This reintroduction will keep promoting a testing culture in schools, knowing that PFC and EN are based on purposes such as controlling, monitoring, certification, and selection, which are not exactly focused on learning improvement (Kellaghan and Madaus, 2003):

We have no consistency in educational policies, despite this minister of education's efforts on consistency. But he has not been able to fight against the higher education lobby, against the private supplementary tutoring lobby for EN preparation. What do we want? Do we want inclusive education for everyone in compulsory school or do we want students to access higher education? (...) Of course, most of these students [referring to low-achievement] will never be able to perform well on EN and, obviously, will not finish secondary education (HT 11).

It should also be noted that the existing PA, PFC, and EN, as other large-scale tests, have as a prime requirement measurement properties amenable to statistical analysis; and reliability and norm-referencing are prime concerns, as they are tests based on psychometric theory (Gipps & Stobart, 2003). Therefore, summoning, in some way, the Stufflebeam's CIPP model (Context, Inputs, Process, and Products – See, e.g., Stufflebeam, 2003:35-36), in PA, PFC, and EN, learning contexts are mostly ignored (accommodations can be made for SEND students), inputs and process are highly controlled and standardized by IAVE; and a major focus is made on products: students tests' results. This fact was also highlighted by some head teachers, referring that learning contexts are not considered when designing external assessment tests, not even contexts are considered in test scores interpretation. Therefore, external assessment tests promote a school ranking based, mainly, on pure external tests' results (Torres et al., 2019), which can be highly unfair for TEIP schools and other schools inserted in less favourable socio-economic contexts that make an excellent work within their community

(Verdasca, 2013). In many ways, for different reasons, the Covid-19 pandemic brought to the forefront the importance of understanding and documenting the contexts of learning and the need to account for them in the design and interpretation of large-scale assessments (NAE, 2021); and we must not neglect them for the future:

In basic education, it is very important to take learning contexts into account, because this constrains learning a lot, particularly, younger students learning (...). Parents' own schooling is reflected in their children's desire to learn, and I think it is critical to consider contexts in PA (HT 10).

For some head teachers, the inexistence of PFC and EN for the conclusion of basic and secondary education would allow higher education institutions to develop their own access model. External assessment in compulsory school, at all grade levels, should become low-stakes, like PA, in line with Fernandes (2019) considerations, referring that we should evolve into another generation of external assessments, based on socially constructed principles, focused on student learning and improvement.

Eventually, for some head teachers, a completely new alternative external assessment model should be developed, without standardized testing. However, it should be noted that large-scale testing seeks to provide a relatively simple and reliable summary of what a student has learned. Alternative large-scale assessments, such as performance assessments, are time and resources consuming, with multidimensional information, where standardization is not possible or even desirable and, as a consequence, reliability, in the traditional sense, is not high (Gipps & Stobart, 2003).

At the end of this discussion, apart from all these different points of view, it is important to reinforce that head teachers agree with external assessment reintroduction in the future. The main divergence comes on its purposes and format.

5.5. Chapter Final Considerations

The lockdowns caused by the Covid-19 pandemic have led to the replacement of large-scale summative assessments with formative assessments built by teachers (Cooper et al., 2022), and the research in this chapter showed that this disruptive scenario sparked a healthy reflection among Portuguese head teachers about external assessment purposes and consequences. This study has also shown that, from the perspective of many interviewed head teachers, external high-stakes

assessments in Portugal have negative consequences on education system functioning, especially, for low-achievement and socio-economic disadvantaged students.

It should be noted that the actors' reactions involved in the education system (head teachers, teachers, pupils, and parents) appear to be strongly conditioned by the context in which standardised assessment systems are developed (Mons, 2009). This was very clear in this chapter research, around a widespread consensus among interviewed head teachers regarding external assessment cancelation during the pandemic, and total discord about its reintroduction in a post-pandemic context.

Note as well that low-stakes assessments, as PA, are perceived by head teachers as an unreliable instrument and, therefore, many of them call for a restructure. Accordingly, it is already known that these tests will be cancelled and *Provas de Monitorização das Aprendizagens* (ModA) for the 4th and 6th grades (See the announcement from the Council of Ministers dated July 18, 2024 – Council of Ministers Announcement, 2024) will be introduced. Future research should investigate the extent to which these new ModA low stake assessment tests are actually able to address the gaps and criticisms that were made to PA.

Finally, it should also be highlighted that there is almost a natural predisposition to associate examinations with high standards, rigor, and a mean to guarantee quality in education systems (Fernandes, 2014b; 2019). Moreover, for many head teachers in Portugal, assessment is still perceived as an instrument to hold stakeholders accountable and to "control" (Santiago et al., 2012). The growing influence of neoliberal and neoconservative agendas in Portugal, and the publicity of schools' excellency based on academic performance of students in EN (Torres et al., 2019), also create additional pressures to keep holding teachers, head teachers, and schools accountable through external assessment tests. Therefore, there is great resistance to change external assessments' purposes, format and stakes, and, in 2024, in a post-pandemic context, EN will be reintroduced with small changes in their weighting to reduce unintended outcomes (see Decree-Law No. 62/2023).

There is evidence that policies to reduce exit exam weighting has some merit, as a means to reduce negative washback of exit exams on teaching and assessment practices. Nevertheless, there is also evidence that reducing weighting alone is insufficient for achieving this goal (Slomp et al., 2020). Future studies in Portugal, beyond 2024, will probably confirm that.

CHAPTER 6

Examining the Effects of the Covid-19 Pandemic and Professional Development on Teachers' Grade Repetition Beliefs and Practice.

Chapter Summary:

Grade repetition has been a controversial educational decision consisting of requiring students to remain at the same grade level for an additional school year. In some countries, during the Covid-19 pandemic, a notable decrease in grade repetition rates was observed, as it has happened in Portugal during the 2019/2020 academic year. The study conducted in this chapter investigates the impact of the pandemic on teachers' beliefs and practices regarding grade repetition, examining the interaction effect of professional development in assessment literacy. Using data from two independent samples collected through two survey questionnaires, we found that while the Covid-19 pandemic had minimal influence on teachers' beliefs regarding grade repetition, professional development had a significant effect.

6.1. Grade Repetition

To deal with the heterogeneity of students' learning and achievement in the classroom, in education systems where the use of grade repetition is a common practice, grade repetition is commonly seen as a *"second chance"* to improve students' knowledge and skills (OECD, 2023b); and grade repetition itself is often seen as *the* educational intervention for low achievement students (Allen et al., 2009). The more teachers consider learning as a mere transmission of knowledge, the more they believe that grade repetition is effective (Santos et al., 2023). However, the use of grade repetition as an educational intervention has been the topic of many controversies. When retained students do not receive extra educational services during the repeated year, studies have shown that grade repetition itself becomes an ineffective tool to deal with underachievement students (Nunes et al. 2018), as it was shown that repeating a school year alone has negative effects on academic achievement, with consequences on socio-emotional and behaviour of students, increasing the likelihood of early school drop (Jimerson, 2001a, 2001b). In education systems with automatic grade promotion, despite the heterogeneity of students' learning and achievement, teachers tend to provide greater support to

improve students' learning (OECD, 2023b), with alternative strategies such as parental involvement, modification of instructional strategies and systematic formative assessment with positive effects on subsequent school achievement (Jimerson, 2001a).

The use of grade repetition is also a common practice to deal with the heterogeneity of students who do not master the autochthonous language, i.e., students with an immigrant background and/or a different mother tongue (European Commission et al., 2020; OECD, 2023). When dealing with cultural and linguistic diversity, Wiese & Nortvedt (2023) showed that grading can affect the trust between student and teacher because students take grades personally, and they argue that teachers should develop a dynamic assessment literacy that includes a broad repertoire of inclusive teaching and assessment. Still, Costa et al. (2015) refer that in Portugal the introduction of changes to previous teaching practices has some resistance. Teachers' concerns to comply the syllabus and their concerns with exams are often obstacles to the development of different teaching strategies and activities to improve learning.

6.1.1. Assessment and Grading During the Covid-19 Pandemic in Portugal

The closure of Portuguese schools during the Covid-19 pandemic unevenly affected students' learning, where the worst affected were pupils from the most vulnerable families and socioeconomic and cultural contexts (Torres & Alves, 2024). Despite relevant institutional and pedagogical responses to deal with the Covid-19 pandemic in Portugal (see, e.g., Flores e Gago, 2020), Portuguese teachers underline the lack of appropriate and clear guidance from the Portuguese government throughout the crisis, especially regarding critical components of the pedagogical process, such as evaluation, assessment, and feedback (Seabra et al., 2021). Additionally, the study conducted by the Conselho Nacional de Educação [National Education Council] (CNE, 2021a) also revealed that the ERT raised on teachers' doubts and questions about assessment methods, instruments, criteria, objects, and trust; where approximately 82% of teachers that participated in their survey indicated that assessment of learning was difficult or even very difficult to perform. Flores et al. (2021) also revealed a tension around assessment and feedback to students' remote homework, with doubts regarding the authorship of these works. Therefore, as in other countries, while more responsibility was placed on teachers to deal with assessment and grading according to students' learning contexts (Cooper et al., 2022), teachers found it very difficult to be in a situation where final grading had to be done by the individual teacher alone, being unclear what should be included in the students' final grading (Sandvik et al., 2023).

Within this disruptive context, despite students' learning losses, grade repetition rates at all grade levels have considerably dropped in Portugal during the 2019/2020, to the lowest grade repetition rate ever recorded in Portugal (see, Figure 7 – page 60). The shift in the use of grade repetition prompted us to explore whether the Covid-19 pandemic affected teachers' beliefs about grade repetition.

6.1.2. Assessment Literacy and Professional Development During the Pandemic

For many years, teachers' assessment literacy was identified with standardised measurement and classroom testing (Pastore & Andrade, 2019). However, in the late decades of the twentieth century, many studies about classroom assessment contributed to the redefinition of this concept, and the extensive review carried out by Black & Wiliam (1998a) became a seminal work playing a vital role in the redefinition of this concept. Nowadays, assessment literacy is conceptualised on how teachers can use information about student learning to teach more effectively responding to students' learning needs (Pastore & Andrade, 2019).

According to Tomchin & Impara (1992), professional development to improve teachers' assessment literacy should encourage teachers to examine their own beliefs about student-teacher responsibilities on grade repetition, providing research evidence and theory on the implications of grade repetition and alternative practices, as well as a schoolwide support structure for developing and implementing effective classroom strategies. Such approach, recognizing the importance of research, theory, and practice, can encourage authentic change with positive outcomes for students. Training programmes aimed at changing teachers' beliefs should encounter more resistance through the process, but the effects, once generated, should be maintained in the long term and have repercussions on teachers' practices (Santos et al., 2023).

In Portugal, aiming to improve Portuguese teachers' assessment literacy and to improve students' learning through teachers' professional development, it was created the national project for Training, Supervision and Research in Classroom Assessment (the MAIA Project). When schools closed due to the Covid-19 pandemic in March 2020, it was decided to continue the MAIA project with distance training (Fernandes, 2021). Therefore, it was decided to test whether the participation in this professional development programme affected Portuguese teachers grade repetition beliefs.

6.1.3. Concepts Clarification for the Study

In the field of educational assessment, a wide diversity of terms has been used to study *"teachers' cognitions"* about grade repetition, such as *attitudes, beliefs, conceptions, knowledge,* and *perspectives* (Barnes et al., 2015; Crahay, 2010; Pajares, 1992). Although attitudes and beliefs are

considered different concepts (e.g. Eagly & Chaiken, 1993; Sudman & Bradburn, 1982), according to Pajares (1992), teachers' attitudes about education – about schooling, teaching, learning, and students – have generally been referred to as teachers' beliefs.

Teachers' beliefs about assessment of learning and, more specifically, about grade repetition, as some studies have shown, are deeply rooted in school culture (Goldring, 2002; Santana, 2019). Thus, beliefs can be understood as ideas or propositions that individuals take to be true (regardless of their scientific validity) and are fundamental in shaping social norms, values, and cultural practices.

Beliefs also play a crucial role in social cohesion, collective identity, and the way individuals interpret and interact with the world around them. This definition is based on the conception that beliefs are transmitted through processes of socialization, communication, and within institutional and relational frameworks. This is very evident when it comes to the socialization and professional development of teachers (Caria, 2000; Flores & Day, 2006). This conceptualization can be further complexified by examining the moderating effect that the educational policies of schools and the country have on the relationship between beliefs and grade repetition practices (Santos & Monteiro, 2023). In this sense, this chapter also provides interpretive perspectives on the influence of public programs or policies, such as the MAIA Project, on grade repetition beliefs and practices, with the possibility of adaptation and recreation by school policies (Fernandes, 2021).

6.2. Method

As referred in section 3.4, to understand how the Covid-19 pandemic affected Portuguese teachers' grade repetition beliefs, Mann-Whitney U tests were used to compare the responses given by teachers in two survey questionnaires, i.e., using two independent samples from two different time periods: Sample A, collected before the Covid-19 pandemic, in April 2019, *and* Sample B, collected in a post-pandemic period, in October 2023.

6.2.1. Measures

In Sample A, to understand the latent structure underlying the set of four items to measure grade repetition beliefs (Tabachnick & Fidell, 2014), and to know the validity of these four items to accurately measure the intended construct (Groves et al., 2004), a factor analysis with varimax rotation was performed (KMO = .68, Bartlett test χ^2 (6) = 1 219.964, *p* <.001), with all items remaining on the same factor, with loadings ranging from .632 to .833. When variables within factors are highly intercorrelated, as in this case, factors are assumed to represent dimensions within the data,

corresponding to concepts that cannot be adequately described by a single measure (Field, 2013; Hair et al. 2006). Therefore, as the four items remained in the same factor, it is assumed that the four items are measuring the same concept, i.e., teachers' grade repetition beliefs. When measuring the internal consistency of these items, since the Cronbach alpha was .64, the mean of inter-item correlation was also calculated (.32), and reliability was ensured (Clark and Watson, 1995). Note that for internal consistency analysis, item 1.1. was inverted, as this item had a negative loading at the factor analysis.

To understand whether the latent structure underlying this set of four variables in sample B was the same as in sample A, a factor analysis was also performed in sample B (KMO = .67, Bartlett test χ^2 (6) = 1420.581, *p* < .001). As in sample A, all items remained at the same factor, with loadings ranging from .617 to .816. This allows us to assume that these four items are measuring the same construct as in Sample A. Since the Cronbach alpha in Sample B was .63, the mean of inter-item correlation was also calculated (.31), and reliability was also assured (Clark and Watson, 1995). Once again, as in sample A, for the internal consistency analysis in sample B, item 1.1. was inverted, as it had a negative loading at the factor analysis.

6.2.2. Procedures and Data Analysis

To allow comparisons between the two independent samples A and B through the use of Mann-Whitney U tests, the two samples were compiled into a single SPSS (2022) file. Before its compilation, items answered with a ten-point Likert scale in Sample A were recoded into a five-point Likert scale as in Sample B.

On Sample B only, items on who is most responsible for students' educational success and grade repetition were analysed through descriptive statistics. Items designed to understand the effect of the MAIA project on current internal assessment practices were analysed using Mann-Whitney U tests.

6.3. Results

6.3.1. Pre and Post Covid-19 Teachers' Grade Repetition Beliefs

When comparing the distributions between sample A *vs* sample B, the period before the Covid-19 pandemic *vs* the period after the Covid-19 pandemic, it was possible to observe that there are no significant differences in teachers' grade repetition beliefs, and significant differences were only found in grade repetition practice, as shown in Figure 25 and Table 21.





Sam ple A, N = 1844, Mean rank = 2210.89
Sam ple B, N = 2655, Mean rank = 2283.01



Sam ple A, N = 1844, Mean rank = 2263.56
Sam ple B, N = 2655, Mean rank = 2245.68



1.4. I use grade repetition as an educational measure.



Sample B, N = 2655, Mean rank = 2214.05

Note: Sample A was collected in 2019, in the pre Covid-19 pandemic. Sample B was collected in 2023, in the post Covid-19 pandemic. Sources: Cipriano & Martins, 2021; Primary data, 2023

(%)

Table 21. Distributions comparisons between samples and subsamples.

	Ν	Mann- Whitney U	Z value	<i>p</i> value	Effect size <i>r</i>
1.1. Grade repetition should only occur by the end of a cycle.					
Sample A vs Sample B	4 506	2 532 929.5	1.923	.055	-
Sample A vs Subsample B ₀	3 539	1 569 301.0	0.226	.821	-
Sample A vs Subsample B ₁	2 811	963 628.5	3.701	< .001	.070
Subsample B_0 vs Subsample B_1	2 662	886 486.0	3.677	< .001	.071
1.2. For some students, grade repetition is inevitable.					
Sample A vs Sample B	4 505	2 433 975.5	-0.479	.632	-
Sample A vs Subsample B ₀	3 540	1 580 845.5	0.597	.550	-
Sample A vs Subsample B ₁	2 809	853 130.0	-1.892	.058	-
Subsample B_0 vs Subsample B_1	2 661	771 981.5	-2.563	.010	050
1.3. Grade repetition is beneficial for students.					
Sample A vs Sample B	4 507	2 521 378.5	1.593	.111	-
Sample A vs Subsample B_0	3 541	1 643 814.5	2.698	.007	.045
Sample A vs Subsample B ₁	2 810	877 564.0	-0.663	.507	-
Subsample B_0 vs Subsample B_1	2 663	764 848.0	-2.990	.003	058
1.4. I use grade repetition as an educational measure.					
Sample A vs Sample B	4 499	2 352 454.5	-2.290	.022	034
Sample A vs Subsample B_0	3 533	1 516 833.0	-1.371	.170	-
Sample A vs Subsample B ₁	2 810	835 621.5	-2.775	.006	052
Subsample B_0 vs Subsample B_1	2 655	783 574.5	-1.743	.081	-

Notes: Sample A was collected pre Covid-19 pandemic. Sample B was collected post Covid-19 pandemic. Subsample B₀ represents teachers that did not participated in the MAIA project. Subsample B₁ represents teachers that participated in the MAIA project. All test can be found in Appendix C. Sources: Cipriano & Martins, 2021; Primary data, 2023

As before the pandemic, most Portuguese teachers (64.9%) tend to reject the idea that grade repetition should only occur at the end of a study cycle (U = 2 532 929.5, p = .055, N = 4 506). Hence, most teachers (71.7%) continue to consider that for some students, grade repetition is inevitable (U = 2 433 975.5, p = .632, N = 4 505). Moreover, despite the fact that 28.7% of Portuguese teachers are not sure if grade repetition brings benefits to students' learning, most teachers (43.1%) still consider that grade repetition is beneficial for students (U = 2 521 378.5, p = .111, N = 4 507).

Even if no significant changes were found in these grade repetition beliefs, significant changes were found in grade repetition practice. In the post-pandemic period, teachers reported to use less grade repetition as an educational measure (U = 2 352 454.5, p = .022, N = 4 499). Still, it should be noted that this is a very small effect size (r = .034; Cohen, 1992).

Note that in the Mann-Whitney U tests reported in Table 21 and Table 22, the sign on the Z value (whether positive or negative) indicates the direction of the difference between the two groups being compared. This means that if the Z value is negative, the first group has higher ranks when compared to the second group. On the contrary, if the Z value is positive, the second group has higher ranks compared to the first group. Therefore, when calculating the effect size *r* through the Z value ($r = \frac{Z}{\sqrt{N}}$, Field, 2013: 227), the value on the effect size *r* represents the effect strength, and the sign (whether positive or negative) provides insight into the direction of the difference between the two groups under comparison.

6.3.2. Understanding the Effect of Participation on the MAIA Project

In the previous section, Table 21 also shows the results of Mann-Whitney U tests between sample A *vs* subsample B₀, sample A *vs* subsample B₁ and subsample B₀ *vs* subsample B₁. These tests have enabled us to understand the effects of participation in the MAIA professional development project on teachers' grade repetition beliefs, as interpreted in the following points.

6.3.2.1. Grade repetition should only occur by the end of a cycle

When testing sample A *vs* sample B and sample A *vs* subsample B₀, it is possible to observe that no significant changes have occurred before and after the Covid-19 pandemic, for teachers who did not participate in the MAIA project. Although, when testing sample A *vs* subsample B₁ and subsample B₀ *vs* subsample B₁, it was possible to observe that significant changes have occurred, as shown in Table 21. This means that for teachers that participated in the MAIA project, they increased the belief that grade repetition should only be used at the end of a cycle.

6.3.2.2. For some students, grade repetition is inevitable.

When testing sample A *vs* sample B, sample A *vs* subsample B₀, and sample A *vs* subsample B₁, it is possible to observe that no significant changes have happened pre- and post-Covid-19, as also shown in Table 21. This means that the Covid-19 period did not have any effect on this belief. Still, when

testing subsample B_0 vs subsample B_1 , it was possible to observe that significant changes have occurred. This means that teachers that did not participated on the MAIA project comparing to the teachers that did participate, believe more that for some students, grade repetition is inevitable.

6.3.2.3. Grade repetition is beneficial for students.

When testing sample A *vs* sample B, and sample A *vs* subsample B₁, it is possible to observe that no significant changes have happened before and after the Covid-19 pandemic for teachers that participated in the MAIA project, as shown in Table 21. However, when testing sample A *vs* subsample B₀ and subsample B₀ *vs* subsample B₁, this belief about grade repetition benefits increased for teachers that did not participate in the MAIA project. Therefore, the Covid-19 pandemic in interaction with the nonparticipation in the MAIA project, increased the belief that grade repetition is beneficial for students.

6.3.2.4. I use grade repetition as an educational measure.

When testing sample A *vs* sample B, it is possible to observe that significant changes have happened, with a decrease on the use of grade repetition. When testing sample A *vs* subsample B_1 (teachers that participated in the MAIA project), this effect is also significant, as shown in Table 21. However, when testing sample A *vs* subsample B_0 , and subsample B_0 *vs* subsample B_1 , no significant changes were found. This means that the use of grade repetition decreased regardless of participation on the MAIA project, but it decreased more for teachers that participated in the MAIA project.

6.3.3. Understanding the Effects of Sex, Age, Qualification, School Sector and Teaching Level on Teachers' Grade Repetition Beliefs

When performing the same Mann-Whitney U tests within subgroups, some significant changes were also found with respect to sex, qualification, school sector and teaching level. While Table 22 reports only the subgroups where the Mann-Whitney U tests showed significant changes, all tests can be found in Appendix E.

Table 22. Distributions comparisons within subgroups

	Ν	Mann- Whitney U	Z value	<i>p</i> value	Effect size <i>r</i>
1.1. Grade repetition should only occur by the end of a cycle.					
Female subgroup					
Sample A vs Sample B	3 381	1 422 429.5	2.208	.027	.038
Sample A vs Subsample B ₀	2 629	869 182.0	0.292	.770	-
Sample A vs Subsample B ₁	2 083	553 247.5	4.208	< .001	.092
Subsample B_0 vs Subsample B_1	2 050	539 110.5	4.146	< .001	.092
Public school subgroup					
Sample A vs Sample B	4 270	2 292 183.0	2.030	.042	.031
Sample A vs Subsample B ₀	3 351	1 409 676.5	0.425	.671	-
Sample A vs Subsample B ₁	2 697	882 506.5	3.587	< .001	.069
Subsample B_0 vs Subsample B_1	2 492	778 355.0	3.354	< .001	.067
1 st cycle (grades 1-4) subgroup					
Sample A vs Sample B	898	112 349.5	3.671	< .001	.123
Sample A vs Subsample B ₀	710	67 638.0	2.153	.031	.081
Sample A vs Subsample B ₁	581	44 711.5	4.449	< .001	.185
Subsample B_0 vs Subsample B_1	505	33 823.0	2.690	.007	.120
1.3. Grade repetition is beneficial for students.					
3 rd cycle / Secondary (grades 7-12) subgroup					
Sample A vs Sample B	2 640	881 374.5	2.317	.020	.045
Sample A vs Subsample B ₀	2 057	569 451.0	3.173	.002	.070
Sample A vs Subsample B ₁	1 647	311 923.5	0.198	.843	-
Subsample B_0 vs Subsample B_1	1 576	268 547.0	-2.496	.013	063
1.4. I use grade repetition as an educational measure.					
Male subgroup					
Sample A vs Sample B	1 071	128 533.0	-2.955	.003	090
Sample A vs Subsample B ₀	873	85 358.5	-1.949	.051	-
Sample A vs Subsample B ₁	711	43 174.5	-3.179	.001	119
Subsample B_0 vs Subsample B_1	558	32 366.0	-1.842	.065	-
ISCED 6 subgroup					
Sample A vs Sample B	3 430	1 365 765.5	-2.101	.036	036
Sample A vs Subsample B ₀	2 695	880 256.0	-1.306	.191	-
Sample A vs Subsample B ₁	2 145	485 509.5	-2.469	.014	053
Subsample B_0 vs Subsample B_1	2 020	454 346.0	-1.458	.145	-
Public school subgroup					
Sample A vs Sample B	4 265	2 124 563.5	-2.240	.025	034
Sample A vs Subsample B ₀	3 346	1 355 132.0	-1.430	.153	-
Sample A vs Subsample B ₁	2 697	769 431.5	-2.556	.011	049
Subsample B ₀ vs Subsample B ₁	2 487	696 108.0	-1.451	.147	-

Notes: Sample A was collected pre-COVID19, in April 2019. Sample B was collected post-Covid19, in October 2023. Subsample B_0 represents teachers that did not participated in the MAIA project. Subsample B_1 represents teachers that participated in the MAIA project. All test can be found in Appendix E. Sources: Cipriano & Martins, 2021; Primary data, 2023 Within these subgroup tests with significant changes reported in Table 22, it seems to be relevant to highlight two of them when considering teaching level: The first cycle teachers' subgroup, where regardless of their participation on the MAIA project, has increased the belief that *grade repetition should only occur by the end of a cycle*; and the 3rd cycle and secondary education teachers' subgroup, where teachers that did not participate in the MAIA project increased the belief that *grade repetition is beneficial for students.* For 3rd cycle and Secondary teachers that participated in the MAIA project, this belief has not changed in the post Covid-19 pandemic period.

6.3.4. Responsibility for Educational Success and Grade Repetition

When it comes to who is responsible for a student's success, whether or not teachers participated in the MAIA project, teachers believe that students are primarily responsible for their own success, followed by the Ministry of Education and its educational policies. Teachers come next, with schools and their internal organization, as well as parents, being considered the least important. On a similar question, when it comes to who is responsible for a student's grade repetition, teachers still believe that students are primarily responsible, followed by the Ministry of Education and its educational policies. Schools and their internal organization come next, while teachers and parents are considered the least responsible. Note that there is a shift on teachers' degree of responsibility.

	Subample B ₀	Subsample B_1
	(relative %)	(relative %)
Educational success depends mostly on:		
Students themselves	46.1	42.4
Educational Policy	21.4	23.9
Teachers	16.7	16.3
School internal organization	10.9	12.5
Parents/Guardians	5.0	4.9
Grade repetition depends mostly on:		
Students themselves	59.2	52.5
Educational Policy	25.0	26.1
School internal organization	7.7	10.1
Teachers	5.0	6.2
Parents/Guardians	3.1	5.2

Table 23. Teachers' Perspectives on Responsibility for Educational Success and Grade Repetition, 2023 (%)

Source: Primary data, 2023

6.3.5. The MAIA Project Effects on Classroom Assessment Practices

To teachers in Sample B, it was also asked to reflect on their practices comparing to the pre-covid period. Comparing to the pre-Covid-19 period, in the post-pandemic context, teachers reported diversifying their assessment instruments more (73.5%), promoting students' independent work and self-regulation of their learning (65.6%), and placing a greater focus on formative assessment to improve students' learning (64.5%), as shown in Figure 26.

Figure 26. Teachers' post-pandemic assessment practices by sub-sample, 2023 (%)



a) Nowadays, I diversify my

Sample B0, N = 1695, Mean rank = 1283.11
Sample B1, N = 965, Mean rank = 1413.74



Sample B0, N = 1689, Mean rank = 1298.13
Sample B1, N = 961, Mean rank = 1373.60

nor agree

b) Nowadays, I promote students' independent work and self-regulation of their learning more.



Sample B0, N = 1695, Mean rank = 1291.38
Sample B1, N = 967, Mean rank = 1401.82





Sample B0, N = 1691, Mean rank = 1337.55
Sample B1, N = 962, Mean rank = 1308.46

Source: Primary data, 2023

Notably, although with a small effect size (Cohen, 1992), for teachers who participated in the MAIA project, these levels are statistically significantly higher than for those who did not participate, as reported in Table 24.

	Ν	Mann- Whitney U	Z value	<i>p</i> value	Effect size <i>r</i>
a) Nowadays, I diversify my assessment instruments more.					
Subsample B_0 vs Subsample B_1	2660	898 161.500	4.467	< .001	.087
b) Nowadays, I promote students' independent work and self- regulation of their learning more.					
Subsample B_0 vs Subsample B_1	2662	887 529.500	3.744	< .001	.073
c) Nowadays, I focus more on formative assessment and improving my students' learning.					
Subsample B_0 vs Subsample B_1	2650	857 786.000	2.537	.011	.049
d) Nowadays, I focus more on summative assessment and grading my students.					
Subsample B_0 vs Subsample B_1	2653	795 533.000	-0.966	.334	-

Table 24. Teachers post-pandemic assessment practices' distributions comparisons between subsamples

Source: Primary data, 2023

Regarding summative assessments and grading, teachers tend to reject the idea that they are more concerned in the post-pandemic period (43.1%), and there are no statistically significant differences between teachers who participated in the MAIA project and those who did not.

6.4. Discussion

Regarding the use of grade repetition as an educational intervention, when comparing Sample A (the pre-Covid-19 period) with Sample B (the post-Covid-19 period), this sub-study revealed that Portuguese teachers reported a decrease in the use of grade repetition, continuing the previous trend towards its reduction (DGEEC, 2023; OECD, 2023b). Although the effect size is very small, it is statistically significant. This finding aligns with Portuguese institutional statistical indicators that reflect this trend over the last decades (DGEEC, 2023), and it seems that the Covid-19 pandemic has had a positive impact on deepening this trend. Grade repetition rates experienced a considerable decrease across all grade levels during the 2019/2020 school year (Cipriano & Martins, 2023), and according to DGEEC data (2023), this trend continued in subsequent years, aligning with the pre-existing downward trajectory (see Figure 7, page 60). A possible cause for this trend reduction might be, as reported by

teachers, the use of a greater diversity of assessment instruments in the post-pandemic context, promoting students' independent work and self-regulation of their learning, with a greater focus on formative assessment to improve students' learning. Yet, further research should be developed on this topic.

When considering Portuguese teachers grade repetition beliefs, no significant changes were found in the two independent samples from the two different time periods. As before the Covid-19 pandemic, although Portuguese teachers are not sure if grade repetition is beneficial to students, they continue to consider that for some students grade repetition is inevitable, and they do not agree that grade repetition should only occur by the end of a cycle. This fact is in line with literature, referring that teachers' beliefs tend to self-perpetuate over time (Pajares, 1992).

Still, despite the great difficulty in making teachers' beliefs evolve (Crahay, 2010), teachers do change. When experiencing voluntary reflective and collaborative professional development programmes with adequate support and follow-up, they can create a dramatic change such as shifts in orientations and beliefs (see, e.g., Le Fevre & Richardson, 2002; Richardson & Placier, 2001). This fact is also supported in this sub-study when analysing data from sample B by subgroups (B₀ and B₁), i.e., considering whether or not teachers participated in the MAIA project. Although with very small effect sizes, significant changes were found in teachers who participated in the MAIA project. Teachers who have participated in the MAIA project have positively changed their grade repetition beliefs. These findings are also in line with other studies in Portugal and Brazil that reveal that teachers with higher qualifications believe less in the benefits of grade repetition as an educational intervention, and teachers' professional development is related to their position regarding grade repetition (Ribeiro et al., 2018; Santos et al., 2023).

It should also be highlighted that in the Post Covid-19 context, 1st cycle teachers increased their belief that grade repetition should happen by the end of a Cycle. This is a relevant finding since 1st cycle teachers are sole-teachers for all subject areas, and it seems that they are finding new forms to deal with the heterogenicity of their classes' achievement other than the use of grade repetition. Additionally, institutional statistical indicators confirm that these are the grade levels with lower grade repetition rates in Portugal (DGEEC, 2023). On the other hand, in the Post Covid-19 context, 3rd cycle and secondary education teachers who did not participate in the MAIA project reported an increasing belief that grade repetition is beneficial for students. This is also a relevant finding, since those are the grade levels with higher grade repetition rates in Portugal (DGECC, 2023). Hypothetically, these teachers tend to struggle more with the heterogeneity of students' achievement, especially in the post Covid-19 context with students' learning losses, and they tend to consider that students should have been retained in a previous grade level. However, further research on this topic should also be developed to confirm these hypothesis.
Note as well that, as suggested by Tomchin and Impara (1992), when examining teachers' own beliefs about student-teacher responsibilities concerning grade repetition, it was found that teachers tend to externalize responsibility. Students are considered primarily responsible for their own (lack of) success, followed by educational policy, while teachers are seen as among the least responsible. These findings align with those of Cipriano & Martins (2021).

6.5. Chapter Final Considerations

As teachers tend to replicate practices from their past experienced as a student, and younger teachers tend to replicate older teachers work, schools are known to play a key role in shaping young teachers' understanding of teaching, including instruction and assessment practices (Caria, 2000; Flores & Day, 2006). Consequently, teachers' beliefs towards grade repetition are formed in an early stage of teachers' career, and they tend to self-perpetuate with regard to reason, time, schooling, or experience (Pajares, 1992). This study showed that they also tend to self-perpetuate to disruptive events, as the Covid-19 pandemic lockdowns and the ERT.

Nevertheless, grade repetition rates drop during the 2019/2020 school year were effective in Portugal, but they seem to be related to doubts and questions about assessment methods, instruments, criteria, objects, and trust (CNE, 2021), rather to a shift on teachers' grade repetition beliefs. Without appropriate and clear guidance for evaluation, assessment, and feedback (Seabra et al., 2021), being unclear what should be included in the students' final grading (Sandvik et al., 2023), teachers used less grade repetition as an educational intervention in the 2019/2020 school year.

However, even with very small effect sizes, this sub-study also showed that teachers' predispositions towards grade repetition can be improved through meaningful, reflective, and collaborative professional development programmes (Le Fevre & Richardson, 2002; Richardson & Placier, 2001), such as the MAIA project in Portugal. This fact highlights the importance for policymakers to consider that teachers and professional development must be the cornerstone of any systemic reform directed at improving schools (Madaus, 1993), which is more effective to change teachers' beliefs than disruptive events such as the Covid-19 pandemic and the ERT.

This page intentionally left blank

CHAPTER 7

Introducing Technologies into National Large-Scale Testing: Are we Ready?

Chapter Summary:

In recent decades, the potential benefits of introducing technologies into large-scale tests have been much discussed. Yet, the path to effective technology use on large-scale testing has fallen short of expectations, especially when these tests have medium or high stakes for students. After a temporary cancelation of external assessment of learning due to the Covid-19 pandemic, the Portuguese government is taking a step forward. With a gradual implementation until 2025, the reintroduction of external assessment of learning in the Portuguese education system foresees all national assessment tests and all national examinations in digital format with the implementation of the *External Assessment Dematerialization Project* (DAVE). However, assessment reforms, such as DAVE, raise concerns within school communities. In this chapter, to identify and analyse these concerns, it was used data collected with the previous 32 semi-structured interviews with head teachers from mainland Portugal. To understand to what extent these concerns were considered and covered in DAVE's design and implementation, a supplementary interview was conducted with the president of the Portuguese institute responsible for DAVE. In addition, it was collected and analysed data from the survey with 2 673 teachers to find out their degree of agreement with the implementation of DAVE. Results show that DAVE raises many concerns among head teachers, and it is not fully accepted by the teachers.

7.1. Large-Scale Testing

Since testing was first introduced in China in 210 BC, it has radically evolved over the centuries with, for example, the introduction of paper-based formats, the introduction of standardized procedures, and the quantification of performance (Madaus, 1993; Madaus & Russell, 2010). The success of testing in schools led to a widespread establishment of large-scale testing in school systems, and large-scale testing is still used as a mean to collect information about the students, schools, and education systems (Fernandes, 2019).

Large-scale testing usually involves the application of a common paper-based test, constructed on prescribed syllabi in curriculum, with emphasis on written tasks, serving several functions, and it is set or controlled by an agency external to the schools from which candidates come, at a state or national

level (Kellagham & Madaus, 2003; OECD, 2023a). Large-scale tests also have a standardized universal design, to maximize accessibility for all intended test takers, knowing that all test takers should have an unobstructed opportunity to demonstrate their standing on the construct being measured. Nevertheless, when the test is not appropriate for all test takers in the intended population, accommodations can be made, retaining the comparability of test scores. To allow comparability of test scores in large-scale testing, the processes by which test takers' responses are evaluated and scored are also standardized (AERA et al., 2014).

Considering the purposes and consequences of large-scale testing, literature often makes a distinction within large-scale tests based on their associated stakes (e.g., Stobart & Eggen, 2012). In the definition of the OECD (2023), national or central assessment tests are standardized student achievement tests, and test results do not have an impact on students' progression through school or certification. On the other hand, national or central examinations are standardized student tests that have a formal consequence for students, such as an impact on a student's eligibility to progress to a higher level of education or complete an officially recognized degree (OECD, 2023a). With or without consequences on students' pathways, large-scale testing has as prime requirement measurement properties amenable to statistical analysis. Therefore, reliability and norm-referencing are prime concerns, as the tests are based on psychometric theory (Gipps & Stobart, 2003).

7.1.1. Introducing Technologies into National Large-Scale Testing

Large-scale testing has been the topic of different tensions and controversies due to stakeholders' different perspectives, conceptions, and assumptions of their role in school systems (e.g., Baker, 2001; Linn, 2001). Although the real pedagogical value of large-scale testing can be questioned, policymakers continue to insist on their use for reasons that are often associated with the assumption that such tests are a credible measure for quality of teaching, quality of learning, and, in general, the quality of education (Fernandes, 2013, 2014b, 2019). In the turn to the 21st century, some researchers argued that large-scale testing should be reinvented. With the massification of the internet use around the world, computer-based testing would emerge as a disruptive technology (Bennett, 2001) and, according to Kerrey & Isakson (2000), paper-based testing would become a "yesterday's testing technology" because it would be inconsistent with what and how students learn.

More than 20 years after the millennial beginning, in most countries in the world, large-scale testing continues to play a significant role in education systems (Fernandes, 2019), with an increasing number of countries performing national or central assessments tests with no stakes for students, reflecting the existing trend towards monitoring standards and collecting diagnostic information to support achievement (OECD, 2023a). Nevertheless, while sophisticated digital learning platforms,

multimedia technologies and wireless communication are transforming what, when and how learning can take place, it seems that transformation in assessment thinking and practice trails behind. In many settings paper-based testing is still seen as the most reliable way to assess educational achievements (Richardson & Clesham, 2021).

According to Alderson & Wall (1993), tests can be powerful determiners of what happens in classrooms, and, as pointed out by the American web-based education commission report written in 2000,

"(...) perhaps, the greatest barrier to innovative teaching is assessment that measures yesterday's learning goals: It is a classic dilemma: tests do a good job of measuring basic skills, which, in turn, influence the teaching of these skills so students can score well on the tests. Testing works well so long as we are testing the right things (...) What will it take to develop tests that truly reflect what students need to learn for the 21st Century? (...) Above all, it will take a focus on the potential of technology to help us better measure the knowledge, competencies, and understandings we value in education." (Kerrey & Isakson, 2000, p. 70)

If, on the one hand, the introduction of technologies into large-scale testing might promote innovation on teaching and classroom practices, on the other hand, the promotion of large-scale testing as an important component for establishing a competitive market in education can be very harmful, and policymakers should consider that the promotion and support of change should happen within the classroom (Black & Wiliam, 1998b). Hence, note that a contrast exists between the large number of countries that have integrated social and emotional skills into their national curricula and their limited inclusion in national/central assessments tests (OECD, 2023a). Further, the emergence of new skills to be assessed with the use of computer-based testing will certainly make the process of test validation a more complex and challenging process for large-scale test developers (Shaw & Crisp, 2015).

Those who advocate the use of computer-based large-scale testing suggest that it might induce positive impacts in education systems and classroom practices. The potential benefits include: a) time, resources and costs savings in the test administration and scoring; b) great score precision with the possibility to monitor teachers' scoring; c) improvements in assessment validity and reliability; d) improvements in test security; e) greater opportunity to evaluate progress over time and to use test results to influence instruction; f) possibilities for new kinds of questions, using multimedia,

simulations and other resources to assess sophisticated learning goals; and g) adaptive assessments based on the questions asked and on students' previous responses (IAVE, 2022; Kerrey & Isakson, 2000; Oldfield et al., 2012). Nevertheless, they also recognize the barriers that a shift from a paperbased testing system to a computer-based testing system can bring, such as: a) concerns about the capability of the technology to assess all subjects and its consequences on the reliability and validity of high-stakes assessments; b) comparability; c) lack of staff, time and training to use computer-based testing on a pedagogical perspective; d) students inequal access to practice on the relevant software or devices; e) cost of investment; f) user verification and security issues; g) lack of suitable physical spaces and devices in schools; h) ineffective and/or unreliable broadband, wi-fi and network capabilities; and j) the consequences on public opinion about the security, fairness and/or malpractice (Ofqual, 2020; Oldfield et al., 2012).

Considering the barriers and negative impacts that computer-based testing can have, it seems that many governments are being cautious when introducing computer-based technology on large-scale testing, especially when they have associated stakes, as examinations have. While the number of OECD countries using computer-based technology in national/central assessment tests has increased from 8, in 2015, to 21 in 2023, computer-based examination is not yet widely implemented. Only eight countries use computer-based uniform technology for at least one exam (OECD, 2023a). These figures might reflect the existing fears and uncertainty about the positive and negative impacts of computer-based testing initiative, Tomas et al. (2015) highlight that research on e-assessment has been dominated by a focus on investigating benefits of use rather than building an understanding of development and implementation. When studying the development and implementation of a computer-based testing initiative, their research showed that subtle interplay exists between assessment stakes, type, stages, and modes that should be considered. Therefore, it seems to be crucial to understand the positive and negative impacts of large-scale computer-based tests, and they should be studied and documented in the development and implementation of a computer and implemented.

7.1.2. Window of Opportunity for Computer-Based Large-Scale Testing in Portugal

As referred before in this chapter, barriers to implement a computer-based testing initiative may include costs of investment to suppress the lack of suitable physical spaces and devices in schools, ineffective and/or unreliable broadband, wi-fi, and network capabilities, as well as lack of training to use computer-based testing on a pedagogical perspective (Ofqual, 2020; Oldfield et al., 2012). These

are costs that many countries cannot afford, including Portugal, without additional lines of investment in their annual educational budget.

Although, to response to the urgent need to foster a strong recovery from the Covid-19 crisis, the European Commission created the *Next Generation EU* instrument of €800 billion. Within the *Next Generation EU*, a window of opportunity was opened, and the *Portuguese Recovery and Resilience Plan* (PRR) foresees reforms and investments of €13.9 billion in grants and €2.7 billion in loans, where 22% of these figures will foster the digital transition, including education (EU, 2020). Within the PRR, the TD-C20-i01 package was designed to promote the digital transition in education. This specific financial package aims to: a) remove obstacles to a quality internet access in schools; b) digital and technological equipment's renewal in school; c) remove limitations to the integrated use of technological and digital equipment; d) remove the shortage of specialized equipment to develop digital skills; and e) overcome the insufficient use of digital educational resources in the teaching-learning process and as well as in assessment processes (RP, 2021; SGEC, 2022). Within this TD-C20-i01 financial package from the PRR, the *External Assessment Dematerialization Project* (DAVE) is funded and implemented in Portugal (IAVE, 2022; SGEC, 2022).

7.1.3. Research Opportunities and Questions

With the research carried out to write this chapter, it was sought to get to know the perspectives of Portuguese head teachers regarding the implementation of DAVE, and the research question that underpins this chapter was first directed to Portuguese head teachers:

RQ4: Are schools ready for the implementation of DAVE? Why?

Additionally, to be able to capture the perspective of different stakeholders, it was formulated two other sub-research question, where RQ4.1 was addressed to the president of IAVE, and RQ4.2 was addressed to teachers:

RQ4.1: Were head teachers concerns covered in DAVE's design? How? **RQ4.2**: To what extent do Portuguese teachers agree with DAVE?

7.2. Methods

As referred to in Section 3.5, in the first part of this research, qualitative data was collected using Instrument A. The 32 semi-structured interviews with head teachers also aimed to understand their concerns regarding the implementation of DAVE. The collected data was analysed using MAXQDA software.

In the second part of this research, qualitative data was collected using Instrument C. The interview with the IAVE president aimed to understand how head teachers' concerns were addressed in DAVE's design and implementation. The collected data was then analysed using MAXQDA software, taking into account the identified head teachers' concerns.

In the third part of this research, quantitative data was collected using Instrument B. The teachers' survey questionnaire aimed to understand teachers' levels of ICT use in the classroom, their degree of agreement about the cancellation of external assessments during the Covid-19 pandemic and their reintroduction in the post-pandemic context, as well as their degree of agreement with DAVE's implementation. Descriptive statistics were used to report the collected data.

7.3. Results

7.3.1. Head Teachers' Concerns Regarding the Implementation of DAVE

A preliminary analysis was conducted, considering the three variables of the documents: i) Type of school cluster, ii) Administrative and governance characteristics, and iii) Region by Nomenclature of Territorial Units II. From the preliminary analysis, it was observed that variations in the content of the interviews do not have a relationship with these three document variables.

Then, performing a relation of codes mapping in MAXQDA, by co-occurrence of codes in the documents, it was identified three head teachers' profiles shown in Figure 27. As it has happened before with Figure 24, the Code Map in Figure 27 displays the similarities of categories in a two-dimensional visualization. The more co-occurrences two codes have, that is, the more similarly they are used in the data, the closer they are placed together on the Code Map. The symbols (circles) sizes used in the Code Map represent the code frequencies. The colours of the codes correspond to the calculated cluster assignments on the map (Rädiker & Kuckartz, 2020, pp. 94-96).



Figure 27. Head Teachers' Concerns and Profiles regarding DAVE

Notes. DAVE – External Assessment Dematerialization Project; EA – External Assessment; HR – Human Resources; ICT – Information and Communications Technology. Source: Primary data, 2023

In green, at the top left of Figure 27, the results suggest that head teachers' major concerns are related to the *implementation and accessibility* of digital assessments. This profile is the biggest one as it has the largest number of coded segments. For these head teachers, computer-based testing is inevitable in the near future, but schools are not ready yet. They consider that there is a huge lack of investment prior to its implementation on internet broadband, computers, and technological infrastructure:

At the moment, I am talking to you at this conference-call, and I am not using the school's internet connection provided by the Ministry of Education. I am talking to you through a private connection. Otherwise, we would not be talking. (HT16)

(...) Do schools support an exam under these conditions? For example, I have 150 students who will take the Portuguese exam. But I don't have 150 computers. The Ministry of Education might say that all students have their own computer. But will all computers work at that time? (HT1)

For the introduction of digital approaches in schools and for the maintenance of the equipment, we need to have a technical team in each school, entirely dedicated to these issues. It cannot be the goodwill of teachers to do this. This is a very voluntary way of dealing with things and it is not professional. (HT26)

Furthermore, the non-existence of digital internal assessment practices, and the lack of teachers' and students' literary to perform digital assessments increases schools' challenges and head teachers' concerns. Therefore, for these head teachers, computer-based large-scale testing, with the whole population of students instead of a sample, might be difficult or even impossible to implement:

In schools, we do not have internal tests with these characteristics. This requires a very specific logistics, and it requires that all schools have adequate resources. (HT15)

(...) a lot of work needs to be done with the teachers. Because students cannot use the computer only at the examination. There must be training and a consolidation of practices that must happen before in the classroom. (HT12)

In younger age groups, particularly in the second grade, it is complicated. We know that our children know how to use tablets and mobile phones, but when they get to a computer, it is not quite the same. (HT8)

We know that PISA is done like this, but it is done with a sample. With a sample, it works. With a school population with big differences between them, at the same time, with the internet working as it works in schools... I don't know.... (HT 25)

The dark blue in Figure 27, indicates head teachers who considered that *Schools are ready!* They consider that there is already a regular use of ICT in the classroom, students have their own technological kit with a personal computer and a personal internet access, and there is already some use of computer-based tests in classrooms. Also, most of these schools have participated in other computer-based large-scale tests such as TIMSS, PISA, and PIRLS, as well as in IAVE's pilots. Therefore, for these head teachers, schools are ready:

We have a particularity that does not occur in many schools: In addition to the [technological] kit that each student received, all students have access to a computer at school. But I know that this is a school that does not exist in many places. (HT 31)

At the ninth-grade level, there are already many teachers who do some kind of assessment through digital resources and digital applications. (HT14)

We, by chance, are a school that a few years ago did the Mathematics PA in digital format. IAVE asked us if we wanted to participate in their pilot, we said yes, and it went beautifully. (HT9)

In light blue, at the top right of Figure 27, illustrates how there are head teachers whose major concerns rely on the *fairness and acceptability* of digital assessments. They recognize that computerbased large-scale testing might promote the digital transition of schools, with benefits on economic and ecological issues, involving fewer human resources for the application and marking processes. However, for these head teachers, the benefits are few considering what is at stake, especially in PFC and EN with implications on the students' pathways:

I understand why and I know that it will ease a lot of things. But I also think that it will bring more entropy than benefits to the external assessment system. At least for the students, who are the ones to whom we should care the most (...). For the students, yes, I think it will make their lives more difficult. (HT24)

Hence, they state that there is a great resistance by teachers and parents for the use of computerbased testing in classroom assessment, and DAVE will promote a stronger inconsistency between classroom assessment and external assessment, with concerns on what and how some subject areas will be digitally tested. Therefore, these head teachers believe that DAVE might increase inequalities between students and schools, and they tend not to agree with DAVE's implementation, as they fear that this new testing system might be unfair and might not be fully accepted by school communities:

I think that if some of our teachers would have to take the exams that our students will have to take on the computer, they would not be able to do it. This is unacceptable. (HT11) When the pandemic came and we had to shut down, without equipment to carry out remote teaching, everyone wanted the government to provide technological equipment for the students. When technological equipment started arriving in schools, parents no longer wanted them. It is weird. (...) Parents do not want to send students to school with technological equipment in their backpacks. Many parents do not want their children to take exams in digital format either. (HT17)

7.3.2. Concerns Covered in DAVE's Design and Implementation

Regarding the head teachers' concerns identified in green on Figure 27, IAVE's president recognize that schools have strong restrictions on internet broadband, computers, and technological infrastructure to support the implementation of DAVE. Conditions are far from being the perfect ones, and planned PRR investments to promote schools' digital transition is still arriving to schools. Nevertheless, the window of opportunity that was opened with the *Next Generation EU* instrument could not be wasted and the Ministry of Education decided to take a step forward: To mitigate internet restrictions in schools, DAVE was designed to allow online and offline test administrations, with a *standalone* solution for schools with poor internet connections. To mitigate the shortage of computers and technological infrastructure, PA were performed with two non-crossing shifts:

The decision to implement DAVE is not only a IAVE's responsibility, but it is also a political decision, and we depend on it. (...) That was a difficult decision to make. From what we have heard from head teachers, there were some schools that were prepared to move forward. In other schools, we knew we would have problems. That is a fact. Actually, I have said it many times: the conditions are not ideal, of course not; but they were considered sufficient to move forward with DAVE's implementation. Therefore, we made the decision with this assumption: Let's find technical solutions so that no school is left without the possibility of taking the tests (...). [When it comes to internet access for computer-based testing], it turned out to be a funny thing: 95% of schools did it online. (President of IAVE)

Regarding the inexistence of digital internal assessment practices, and the lack of teachers' and students' literacy to perform digital assessments, IAVE's president refers that they are very aware of this reality and, therefore, every step is being taken very carefully. Consequently, IAVE is not implementing a truly *born digital external assessments*, but, in general, they are migrating traditional paper-based tests to computer-based tests:

We know that we should only have external assessment using digital supports when the use of digital technology is already a common practice in the classroom. Which is not. What is happening at the moment, we are migrating to a paper-free external assessment system, with the use of digital platforms, without a widespread use of digital in the classroom. It's a fact. However, knowing that, IAVE is being very careful with this transition. There is a sea of possibilities for building items in digital format, such as the use of interactive items, simulation items, etc. We did not want to introduce these things immediately. For several reasons: the first reason, and the most important one, is the fact that digital is not yet an integral part of teaching in our classrooms. It is not. Didactics, in general, are still based on traditional resources. (...) Therefore, this year's standardized tests, the digital PA, were constructed in a very similar way than before, with a structure very similar to the structure of a paper-based test. (...) Another reason for this smooth transition, was also our training: We, ourselves, IAVE, we have to evolve, our teams have to evolve, they have to learn how to build tests in digital support, which is not exactly the same thing. (President of IAVE)

When it comes to stakeholders' perceptions about DAVE,

(...) meetings held with head teachers, teachers and parents, everyone understood that large-scale digital testing will cause a drastic reduction in certain bureaucracies. Teachers understand this perfectly. In the marking process, for example, something that used to take two weeks to do, they can do it in half of the time, with many advantages in terms of reliability due to the process being online with the possibility of multiple teachers marking a given item (...). Still, we couldn't make everything go flawlessly and, naturally, there was also criticism.⁴ (President of IAVE)

7.3.3. Teachers' ICT Use in the Post Pandemic Period

On the teachers' survey, it was possible to observe that only half of Portuguese teachers (52.7%) make a regular use of ICT in the classroom, such as the use of internet pages, PowerPoint©, videos, learning platforms and tools for the creation of digital educational resources such as Padlet©, Kahoot©, Quizizz© and Piktochart©.

⁴ See, e.g., ANPRI report (2023) – *Associação Nacional de Professores de Informática* [National Association of Informatic Teachers] – available in Portuguese through https://www.anpri.pt/mod/forum/discuss.php?d=13742



Figure 28. Teachers use of ICT in their classroom practices (%)

Source: Primary data, 2023

Additionally, when compared to their own practices before the pandemic, teachers tend to agree or strongly agree that nowadays they teach with online support for students who are sick and cannot come to school, they communicate and provide feedback to their students through email or digital platforms, and they use video conference calls to communicate or meet with colleagues, students, and/or parents, as shown in Figure 29.



Figure 29. Degree of Agreement (%) with Digital Changes Due to the Covid-19 Pandemic



Nowadays, I communicate and provide feedback to my students through email or digital platforms.

Nowadays, I use video conference calls to communicate or meet with colleagues, students, and/or parents.



Yet, when it comes to classroom testing, as shown in Figure 30, 10.8% of teachers never use paperbased tests while 58.1% have a regular use of paper-based tests. As for digital testing, 27.3% of teachers never use digital based tests in classroom assessment and only 19.1% have a regular use of this kind of testing.





Source: Primary data, 2023

7.3.4. Teachers' Acceptability of External Assessments

To understand teachers' acceptance of DAVE, it is necessary to first consider their stance on the cancellation of external assessments during the Covid-19 pandemic and their reintroduction in the post-pandemic period. Generally, teachers believe that cancelling these tests during the pandemic was a good decision, particularly for the 2nd grade PA (77.4%), the 5th grade PA (70.5%), and the 8th grade PA (68.0%). The cancellation of the PFC and EN is more divisive, with 51.3% agreeing with the cancellation of the Portuguese and Mathematics PFC during the Covid-19 pandemic, while only 49% support the cancellation of the EN for the conclusion of secondary education.

Regarding external assessments reintroduction in the post-pandemic, 51.6% of teachers consider that PA should not be reintroduced at all in the education system. Among the teachers who agree with at least one PA (i.e., within the subgroup of teachers who participated in the survey and agree with the reintroduction of, at least, one PA (n=1 293; 48.4%)), they tend to support the existence of these PA in the 2nd and 3rd Cycle of basic education, while rejecting them in the 1st cycle and in secondary education, as depicted in Figure 31.



Figure 31. Agreement with PA existence (%) within the subgroup of teachers who agree with, at least, one PA

Regarding exams with stake for students, only 14.4% of surveyed teachers consider that exams should not exist at all. Among the teachers who agree with at least one exam (i.e., within the subgroup of teachers who participated in the survey and agree with the existence of exams (n= 2 287; 85.6%)), they tend to support their existence in the 3rd Cycle of basic education and in Secondary Education. Yet, they completely reject them in the 1st and 2nd cycle of basic education, as depicted in Figure 32.

Source: Primary data, 2023

Figure 32. Agreement with exams existence (%) within the subgroup of teachers who agree with, at least, one exam



Source: Primary data, 2023

When it comes to the purposes why exams should exist on secondary education, teachers mostly support their existence for higher education access, as Table 25 shows.

Tuble 25. Teachers positioning regurating the purposes of EN in Secondary Education (n, %)				
	п	%		
EN should exist only for access to higher education.	892	33.4		
EN should exist only for the conclusion of secondary education, and higher education institutions should develop their own selection and admission models.	454	17.0		
EN should exist both for the conclusion of secondary education and for access to higher education.	789	29.5		
EN should be completely abolished, and higher education institutions should develop their own selection and admission models.	538	20.1		

Table 25. Teachers positioning regarding the purposes of EN in Secondary Education (n, %)

Source: Primary data, 2023

7.3.5. Teachers' Acceptability of DAVE

When asked if they agree with DAVE, a major part of Portuguese teachers agree with, at least, one large-scale test to be performed on the computer (57.0%), while 43.0% of teachers do not agree with any kind of computer-based large-scale test.

Nevertheless, when considering the whole sample and which test(s) should be computer-based, teachers tend to reject the administration of the second, fifth and eighth grades PA on the computer (only 38.5%, 35.7% and 38.8%, respectively, agree with computer-based PA; as shown in Table 26).

Hence, only 29.4% agree with 9th grade computer-based PFC, while 23.6% agree with 11th and 12th grades computer-based EN.

Among the teachers who agree with at least one computer-based test (i.e., within the subgroup of teachers who participated in the survey and agree with some test(s) being performed on a computer (n= 1524; 57.0%)), they tend to support the existence of computer-based PA, with no stakes for students, as also shown in Table 26. In addition, within this subgroup of teachers that agree with some computer-based test(s), they still tend to support the existence of computer-based PFC in the ninth grade, which already involves some stakes for the conclusion of basic education. Although, these teachers tend to reject computer-based secondary education final exams (EN), which are used for the conclusion of secondary education and higher education access. This means that within teachers who agree with some computer-based test(s), they tend to reject computer-based test(s), they tend to reject computer-based test(s) as stakes are getting higher.

		Sample	Subgroup
	п	%	relative %
 Agree with 2 nd grade computer-based PA	1 030	38.5	67.6
Agree with 5 th grade computer-based PA	954	35.7	62.6
Agree with 8 th grade computer-based PA	1 038	38.8	68.1
Agree with 9 th grade computer-based PFC	787	29.4	51.6
Agree with 11^{th} and 12^{th} grades computer-based EN	630	23.6	41.3

Table 26. Degree of Agreement with DAVE on the Sample and Subgroup (n, %, relative %)

Notes. On the sample (with n = 2673), a group of teachers (with n = 1524) agree with some kind of computer-based testing. On Table 17 it is reported the percentage when considering the whole sample; and the relative percentage when considering this subgroup of teachers only, i.e., within the group of teachers who agree with some kind of computer-based testing. Source: Primary data, 2023

7.4. Discussion

The results showed that, on the one hand, some teachers and head teachers in Portugal consider that the schools are ready, and they agree with the implementation of DAVE. On the other hand, survey data showed that Portuguese teachers tend to reject the existence of PA, as they tend to reject the existence of computer-based large-scale tests. Further, many head teachers consider the digital transition is inevitable in the future, but schools are not ready yet for such a process due to a substantial lack of investment before its implementation. First, concerns about schools' technological capacity to implement such a project arise as PRR investments have not yet arrived to schools. In the TALIS 2018 report, more than 60% of Portuguese lower secondary head teachers have already reported that the school's capacity to provide quality instruction is hindered "quite a bit" or "a lot" by the shortage or inadequacy of digital technology for instruction. Additionally, 35% also reported an insufficient internet access in schools (OECD, 2022). Although the Covid-19 pandemic has brought some technological improvements to schools, most of the interviewed head teachers reported that there is still an enormous lack of equipment and internet quality for such implementation as DAVE. Despite these restrictions, IAVE was able to successfully implement assessment tests (PA) in digital format in 2023, with different solutions being provided in tests design and administration to mitigate these limitations.

Second, many head teachers also consider a general lack of teachers' digital literacy to implement DAVE and, still according to the OECD (2022), less than 60% of Portuguese teachers regularly use ICT for teaching. These figures were also confirmed with our survey, where only 52.7% of teachers reported regular use of ICT for teaching. Note that the average age of teachers in Portugal is quite high, where more than half of Portuguese teachers are over 50 years old (DGEEC, 2023). With the absence of a concerted and continuous training plan for the application of digital educational resources in teaching (CNE, 2021a), and despite the foreseen investments on teachers' capacitation within the PRR (RP, 2021; SGEC, 2022), schools' realities raise doubts among Portuguese head teachers about the gap widening between internal and external assessment practices.

Third, those limitations create fears of social injustice with DAVE, where students will be taught with different or inexistent digital approaches at schools. One cause of social injustice, referring to events before an assessment is taken, is differential opportunity to learn. If some students have not been taught the content being assessed, or have been taught it poorly, they will be disadvantaged in the assessment, compared to other students who have been taught the relevant content well (Nisbet & Shaw, 2022). Also, when technology is involved, it is important that examinees have had similar prior exposure to the technology and that the equipment provided to all test takers be of similar processing speed and provide similar clarity and size for images and other media (AERA et al., 2014). As referred by IAVE's president, to mitigate these limitations, there has been a commitment with schools to provide computer-based tests with very similar characteristics than previous paper-based tests; as well as efforts to involve schools in this external assessment reform, knowing that this process can be jeopardized if educational communities are not involved and engaged with it (Geijsel et al., 2001). In addition, the IAVE president recognized that, considering the Portuguese context, it is not possible nor desirable to go fully paper-free on the different subjects, types, stages, and modes of large-scale tests. As underlined by Tomas et al. (2015), some tests can have hybrid processes, with both paper and

electronic modes within very different parts of the development and implementation processes, as it will happen in Portugal with DAVE.

Finally, many concerns about the fairness of DAVE still arise, especially when external assessments have stakes associated, as in the case of ninth grade and secondary education examinations. All the above-mentioned limitations might conduct to inequities among students and schools (Ercikan et al., 2018); and inequities will impact students' performance. With national assessment tests (PAs), with no stakes for the students, teachers can interpret the test report according to what they know about students and consider that formative information to intervene in the classroom. However, with PFC and EN, exams performed for classification, certification, and selection purposes, this does not happen. Ultimately, inequities will impact decisions based on these assessments, widening the gap between socially advantaged and disadvantaged students (NAE, 2020).

7.5. Chapter Final Considerations

The implementation of a computer-based large-scale testing initiative, such as DAVE in Portugal, brings different types of concerns to school communities. This sub-study showed that, for many Portuguese head teachers, concerns arise about the implementation process and the technological infrastructure of schools. The implementation of DAVE also unveils concerns about the tests' validity, reliability, and fairness. Nonetheless, it should be once again noted that IAVE was successfully able to provide solutions during the implementation of national PA in 2023, and was also able to mitigate identified limitations and head teachers' concerns. Even so, policymakers have to consider that teachers, not assessments, must be the cornerstone of any systemic reform directed at improving schools (Madaus, 1993). To successfully integrate automated technologies in high-stakes contexts, it is required more than guaranteeing technological infrastructure and tests validity, reliability, and fairness. High levels of public trust and acceptability are also needed (Aloisi, 2023). Despite the efforts of IAVE and its capacity to provide different technical solutions on digital PA, our teachers' survey, carried out after the implementation of PA in digital format, showed that DAVE is not fully accepted, especially in high-stakes contexts.

The process of evaluating acceptability of a testing initiative should consider specific circumstances that affect stakeholders' individual standards, and several other aspects, such as applicable laws, regulations, and alternative measurement devices that are readily available (AERA et al., 2014). Portuguese teachers and head teachers have their own conceptions about external assessment purposes and practice (e.g., Cipriano & Martins, 2021; Machado et al., 2022), and this substudy showed that, for these stakeholders, there are additional powerful contextual circumstances

affecting DAVE's trust and acceptability. Regarding applicable laws and regulations affecting acceptability, the debate about trust and acceptability on the use of computer-based technologies when stakes are involved is not new nor restricted to large-scale testing. For example, when it comes to trust and acceptability of e-voting, some countries have completely rejected the use of the internet on voting, while only a few countries have run a significant number of internet elections on a national scale. Hence, of these few countries, only Estonia has continued with plans for universal adoption (Gibson et al., 2016). While new laws and regulations can promote and support computer-based large-scale testing initiatives, decisions about the use and format of external assessment tests are always linked to the political and ideological choices of those who have the responsibility to put them into practice (Fernandes, 2019). In Portugal, it is a political decision.

Finally, policy decisions have technical implications, and technical decisions have policy implications (Madaus, 1993). This sub-study has also showed that DAVE implementation is not free of tensions and criticism concerning policy and technical issues. Technical mechanisms could be found to successfully mitigate all the previous identified limitations in the implementation process of a computer-based large-scale testing initiative. Yet this sub-study also shows that these technical solutions are not enough to provide stakeholders' trust and acceptability regarding this external assessment policy. As it has happened with the algorithm experience in the UK during the Covid-19 pandemic, where massive criticism by students and parents condemned the acceptability and the use of statistical models implemented at the time (Nisbet & Shaw, 2022), the implementation of DAVE might also have a throwback. In the event of large-scale social criticism, on the natural political alternation that characterize democracies, new governments will know that the traditional "paperbased yesterday's testing technology" (Kerrey & Isakson, 2000) is still readily available as a working alternative measurement device (AERA et al., 2014) that is socially accepted. Therefore, it seems that providing public trust and acceptability, especially in high-stakes contexts, will be one of the greatest challenges that large-scale test developers and policymakers will have when introducing technologies into national large-scale testing, knowing that the use and format of large-scale tests for the future, paper- or computer-based, will always be a political decision.

This page intentionally left blank

Conclusion

The work carried out in this thesis attempts to study the relationship between student assessment policy and practice in a highly disruptive context, such as the Covid-19 pandemic. More specifically, considering the pre-existing assessment system, it examines how Decree-Law No. 14-G/2020 was implemented, interpreted, and experienced by Portuguese teachers and head teachers, as well as the effects of this emergency policy on student assessment practices and its implications for the future.

* * *

Summary of Achievements:

In Chapter 1, the theoretical aspects of evaluation and assessment were explored, along with how the Portuguese assessment system is structured in compulsory education. This chapter aimed to provide a framework for understanding how student assessment is conducted in Portugal, including the most recent developments in student assessment policy and relevant institutional projects.

In Chapter 2, some of the major and visible impacts of ERT on student assessment practices and outcomes during the first wave of the Covid-19 pandemic were examined. Through the analysis of statistical indicators and studies conducted during this period, the main implications of Decree-Law No. 14-G/2020 during this initial phase of the Covid-19 pandemic were explored, setting the groundwork for further development in the following chapters of this thesis. Thus, the thesis then developed four analytical dimensions:

- 1. Internal Assessment Practices in ERT;
- 2. Cancellation of External Assessments;
- 3. Grade Repetition Beliefs;
- 4. DAVE's Implementation.

Regarding internal assessment practices in ERT, Chapter 2 initially investigated the sudden implementation of ERT in 2020, which profoundly altered the context and pedagogical dynamics between teachers and students. Simultaneously, the enactment of Decree-Law No. 14-G/2020 introduced additional regulations to internal assessment practices. This new reality impacted teachers'

adaptation and their perceived workloads, topics that were further explored in Chapter 4. Accordingly, Chapter 4 revealed that the relationship between teachers' adaptation to ERT and their perceived workloads was mediated by instructional and assessment practices. Furthermore, it demonstrated that this indirect effect was moderated by adjustments made in assessment methodologies. Therefore, adapting assessment methodologies, as proposed by Decree-Law No. 14-G/2020, emerged as a crucial strategy for facilitating better adjustment to ERT. Moreover, when considering special education teachers exclusively, it was observed that changing assessment methodologies did not moderate the relationship between teachers' adaptation to ERT and instructional and assessment practices. This finding indicates that the difficulties experienced by special education teachers extended beyond the scope of changing assessment methodologies. The findings of Chapter 4 also emphasize the importance of policy-makers to consider learning contexts when defining classroom assessment policies, and the need to provide measures and mechanisms that genuinely ensure inclusion and access to education for all students, allowing their full participation in assessment processes, even in disruptive emergency contexts such as the Covid-19 lockdowns.

Concerning the cancellation of external assessments, the second topic covered in Chapter 2, the relationship between internal and external assessments in light of Decree-Law No. 14-G/2020 was explored. This change had consequences for the conditions under which basic and secondary education are concluded, with implications for access to higher education. Additionally, it was noted that this period served as a catalyst for evaluating higher education access systems, assessing their effectiveness, suitability, and contribution to a more inclusive higher education system. In Chapter 5, this topic was further developed, revealing the perspectives of Portuguese head teachers on the impact of the temporary suspension of external assessments during the Covid-19 pandemic and its implications for school management. From their viewpoint, it was found that the pandemic prompted a reconsideration of the purposes and format of external assessments, highlighting varying perspectives on their reintroduction in the Portuguese context. Furthermore, this chapter examined the negative effects of external assessments, particularly on low-achieving students and school management.

Regarding grade repetition beliefs, Chapter 2 highlighted the significant reduction in grade repetition rates during the 2019/2020 school year. Despite the learning losses caused by the Covid-19 pandemic, a notable decrease in grade repetition rates was observed in Portugal. To understand how the Covid-19 pandemic affected teachers' beliefs about grade repetition and whether the decline in grade repetition rates was accompanied by a change in these beliefs, Chapter 6 examined the impact of the pandemic on Portuguese teachers' views on grade repetition. The sub-study conducted in this chapter revealed that teachers tend to externalize responsibility for students' success and students' grade repetition. Moreover, the Covid-19 pandemic had minimal influence on teachers' beliefs about

grade repetition. However, participation in the MAIA project significantly shifted these beliefs. These findings suggest that teachers' beliefs about grade repetition tend to be self-perpetuating, highlighting the importance of professional development in enhancing assessment practices.

Finally, concerning DAVE's implementation, Chapter 2 identified a lack of teachers' digital competencies for conducting ERT and limited integration of ICT in classrooms before the Covid-19 pandemic. Chapter 7 further explores this topic by examining the implementation of a computer-based large-scale testing system in the post-pandemic period. The results revealed divergent opinions among Portuguese teachers and head teachers. While some believed that schools are prepared for DAVE's implementation and supported it, a major proportion of head teachers expressed concerns about the lack of investment prior to implementation and noted widespread teacher resistance. Consequently, this new testing system presents significant challenges regarding its fairness and acceptability within educational communities.

* * *

Contributions to Student Assessment Policy

Since the establishment of democracy in Portugal in 1974, public policies for student assessment have been defined by political decision-makers in government, often designed and developed with the consultation and support of various stakeholders, organizations, and associations, as well as individual experts and researchers in education. Throughout the 50 years of Portuguese democracy, policies for student assessment have, therefore, set principles, norms, and frameworks for the actions of the various educational agents within the education system, grounded in scientific evidence and ideological philosophies of the governments that implement them.

In 2020, the Covid-19 pandemic created an exceptional emergency context throughout the country, unprecedented in the relatively young Portuguese democracy. This event was so disruptive to the Portuguese society that it was required the implementation of numerous emergency policies at various levels, such as in healthcare, housing, culture, justice, internal administration, and foreign affairs (see, e.g., Carmo et al. 2020; Lomba, 2020). Many emergency policies were also established in the field of education (see, e.g., Costa & Baptista, 2023; Flores & Gago, 2020), including those related to student assessment (see, e.g., Cipriano & Martins, 2023; CNE, 2021a; Decree-Law No. 14-G/2020). Consequently, it is important to emphasize that the changes to student assessment policies in 2020 and the following years of the Covid-19 pandemic did not emerge from a scientific, ideological, or governmental shift, but it emerged from the need to produce legislation to address a disruptive

emergency context. This shift in the driver that triggered changes in student assessment policies is very relevant to be considered, as it highlights the importance for policymakers to recognize the weight and the importance of contexts in which they operate, whether under normal conditions or during emergency situations.

It should be also noted that emergency policies, by definition, are aimed at responding to emergency contexts. As Boin & Hart (2003) pointed out, this can only be achieved if leaders flexibly adapt the policy-making structures and modus operandi of public organizations to the high-pressure crisis environment. Hence, it is also important to underline that it is not easy to determine what must change and what should remain the same. Therefore, leaders must have a clear understanding of what is worth preserving in their society, policy field, or organization during an emergency crisis (Boin & Hart, 2003). In Portugal, since the approval of the Education Act (LBSE) in 1986, which defines the functioning of the Portuguese education system, it had never been necessary to produce emergency policies in education on such a scale. Pondering what should be changed and what should be maintained during the Covid-19 crisis, while seeking to minimize the pandemic's impact on students' opportunities to learn and inequalities affecting their educational trajectories (Martins, 2020), Decree-Law 14-G/2020 was enacted to regulate student assessment policy and practice in this context. This legislation, therefore, introduced a shift in student assessment practices, altering them within educational organizations and among individuals, spanning from IAVE's standardized tests to teachers' individual classroom formative assessment practices (Cipriano & Martins, 2023).

This thesis demonstrates that, at the teaching level, the pandemic and the ERT period introduced significant changes in teachers' practices, many of which persist in the post-pandemic period (Costa & Baptista, 2023). Additionally, when considering the analytical model proposed in Figure 1 (Page 2), this thesis highlights that the changes introduced in internal assessment practices during the pandemic, prompted by the ERT context and the new legal framework for student assessment, contradicted deeply rooted beliefs within the Portuguese teaching profession, particularly regarding the use of assessments for accountability, student selection, and grade repetition (Cipriano & Martins, 2021). Furthermore, this thesis shows that changes in grade repetition and assessment practices during this period were also accompanied by the MAIA project, a professional development program in classroom assessment, which positively influenced teachers' attitudes towards the use of assessments to improve student learning. Concurrently, the thesis demonstrates that while the Covid-19 pandemic contributed to a reduction in grade repetition rates, only the MAIA project promoted a positive shift in teachers' beliefs regarding the benefits of grade repetition for students' learning and educational pathways. This fact highlights the importance of professional development programmes to improve teachers assessment literacy and practices, even when they face some resistance by the teachers.

174

Note that crisis periods may be windows of opportunity to push through reform packages that would be unimaginable during normal times (Boin & Hart, 2003), and the cancellation of external assessment for the conclusion of basic and secondary education appears to have been one of these unimaginable events. As such, during the pandemic, the cancellation of external assessments for the conclusion of basic and secondary education impacted the organization of internal assessments in schools, and this thesis shows that the shift in this paradigm, even if temporary, sparked a healthy discussion about the purposes of external assessment of learning. It also shows how external assessments have a significant impact on the daily management of Portuguese schools (Machado et al., 2022; Torres et al., 2019), and that their cancellation had positive effects by mitigating some of the negative aspects that external assessments impose on the functioning of the education systems.

The analysis carried out in this thesis also shows that emergency policies related to student assessment during the Covid-19 pandemic seemed to have garnered broad consensus during the emergency itself, but they do not enjoy the same consensus in a post-pandemic context. This highlights that the reactions of the actors involved in the education system (head teachers, teachers, pupils, and parents) are strongly influenced by the context in which assessment systems operate (Mons, 2009). It should also be noted that, during the emergency, student assessment policies primarily focused on identifying problems and strategies for their resolution, placing formative assessment – whether internal or external – at the centre of student assessment practices, in an unprecedented manner in Portugal. In a post-pandemic context, however, accountability and selection mechanisms have regained momentum, marked by the reintroduction of external assessment tests for the conclusion of basic and secondary education. In an education system that aims to be compulsory and inclusive, where assessment should fundamentally focus on improving learning, it is crucial to reflect on the mechanisms through which this improvement is effectively achieved. Furthermore, if the main purpose of assessment is to serve the learning goals of education, the reintroduction of external assessments for accountability purposes must be discussed about its relationship with learning (Baird et al., 2017) and to what extent they effectively support these goals.

Within this thesis, it is also emphasized that the digital transition in education is critical, while at the same time exposing the financial, technical, and social vulnerabilities involved in implementing this transition. These vulnerabilities were evident in the challenges faced by teachers and students during the Emergency Remote Teaching period (Flores et al., 2023), as well as in the conditions for implementing the DAVE project (Cipriano & Martins, 2024). Furthermore, the critical problem of the digital transition in education is accompanied by intense discussions within school and scientific communities about the cost-benefit analysis of using technology in educational contexts, and its positive and negative impacts on students' learning (see, for example, Bastani et al., 2024; Jensen et al., 2024; Rogne et al., 2024; Swiecki et al., 2022).

It is also important to note that the post-pandemic period is marked by a political shift in the Portuguese government. With this political change, amendments have already been made to the scientific evidence and ideological principles favored by the new government, which is now responsible for implementing student assessment policies. Consequently, at the level of external assessments of learning, modifications have been introduced regarding low-stake external assessment tests and the grade levels at which they are administered, including changes to the implementation and calendar of the DAVE project and its application when stakes are involved. In addition to these measures concerning external assessment tests, at the internal assessment level, the MAIA project has been canceled, and revisions to the laws that establish the norms and principles for internal assessment practices are currently under review. Nevertheless, until now, it remains unclear how these changes to student assessment policy relate to other levels of assessment within the educational system particularly with respect to the evaluation of the system itself, school evaluations, and teacher performance appraisals – and how they effectively contribute to improving student learning. This lack of specification perpetuates concerns about how these different assessments within the education system align with principles that ensure the elements are complementary and work together (Berman et al., 2020; Colardaci, 2002; Santiago et al., 2012), with the PASEO and the improvement of student learning at the core of the education system.

Thus, finally, this thesis demonstrates that student assessment policies are in constant flux, driven by ideological philosophies, scientific evidence, and contextual factors, significantly impacting the practices of various educational agents operating within the education system, and affecting students' learning and educational pathways. With or without emergency contexts, it seems that a struggle will always exist in finding a consensual approach on how to maximize <u>all</u> students' learning.

* * *

Limits and Challenges for Research

The aim of any research is to produce valid and reliable conclusions. Valid conclusions refer to the extent to which an empirical measure accurately reflects the true meaning of the concept being studied (Babbie, 2013: 191). It concerns the inferences drawn from the data, i.e., it concerns with the integrity of the conclusions that are generated from a piece of research (Bryman 2012: 47). Reliable conclusions, on the other hand, pertains to whether a particular technique, when applied repeatedly to the same object, yields consistent results each time (Babbie, 2013: 188). It concerns with the question of whether the results of a study are repeatable (Bryman, 2012: 46).

A primary limitation concerning the validity of this study is related to understanding what actual assessment practices were performed by teachers during ERT. Additionally, it would have been important to observe how the cancellation of external assessments impacted teachers' internal assessment practices across different subject areas. Throughout the study, the analyses carried out relied on self-reports from teachers and head teachers, collected through surveys or interviews, which may not fully reflect their actual experiences and practices. Observing and documenting real assessment practices during ERT would have been valuable.

A second limitation concerning the validity of this study, related to the first, involves understanding the extent to which the alleged increase in formative assessment practices during ERT was effective and its effects on mitigating students' learning losses. Additionally, it would have been important to know how the feedback provided by teachers during ERT was received and appropriated by students, as well as how this period effectively enhanced students' self-regulation processes to mitigate their learning losses.

A third limitation of this study, concerning the reliability of the conclusions, pertains to the replicability of the study. As time goes by, the memories that teachers and head teachers have regarding their lived experiences in ERT tend to lose detail or fade away, making it difficult or impossible to replicate the study in the future, as the object of study tends to disappear. In this regard, the object under study, it is important to note that this thesis refers to an emergency context that no longer exists, making it factually impossible to observe or document effective assessment practices during the Covid-19 lockdowns, as it is not possible to re-experience or recreate the ERT period. However, almost from an anthropological perspective, I should mention that I personally experienced ERT during the first lockdown as a music education teacher, along with the difficulties and anxieties my fellow teachers and I faced in ensuring our school continued to function virtually for our students. This was a unique experience and a memory that I, along with all teachers in Portugal in their local contexts, will carry throughout our careers. Notably, the experiences reported by teachers and head teachers in this thesis regarding their assessment practices during ERT closely align with my own experiences as a teacher at my school during that time.

While it is impossible to re-experience or recreate the ERT period to observe and document effective practices for understanding teachers' assessment approaches during ERT, further research on current assessment practices and students' self-regulation processes could and should be developed in the post-pandemic period, exploring their potential links to the Covid-19 lockdowns and the ERT period.

* * *

Final Considerations:

The thesis *"Student Assessment in Portugal: Impacts of the Covid-19 Pandemic in an Ongoing Process of Change"* aimed to provide a comprehensive understanding of the implications of the Covid-19 pandemic on student assessment in Portugal. Emphasizing that student assessment serves as a means of gathering information to inform decision-making processes across different levels of education systems, this thesis elucidates how the pandemic altered both the processes of information gathering and the subsequent decisions made within the education system, spanning from policy to practice. Furthermore, this thesis acknowledges the dynamic nature of student assessment, which requires continuous adjustments to evolving contexts. It also recognises that the challenges faced by education systems are inherently complex, involving numerous stakeholders, variables, perspectives, and tensions. Therefore, broad consensus is difficult to achieve.

Lastly, it is imperative to acknowledge that there is no return to a pre-pandemic "normal." In the post-pandemic period, old and new challenges in student assessment coexist, such as the real and effective integration of classroom assessment as a pedagogical process to improve students' learning, the impacts of teacher shortages on students' opportunities to learn and the fairness of their high stakes assessments, or the integration of Artificial Intelligence into instruction, assessment, and curriculum management. These old and emerging challenges once again underscore the dynamic nature of the field of educational assessment and highlight the crucial role of public policies in regulating these processes within education systems.

References

Abedi, J. (2009). Computer Testing as a Form of Accommodation for English Language Learners, *Educational Assessment*, 14:3-4, 195-211. http://dx.doi.org/10.1080/10627190903448851

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park: Sage. Alderson, J.C. and Wall, D. 1993: Does washback exist? *Applied Linguistics* 14, 115-29.

- Aldrich, J. O. & Cunningham, J. B. (2016). Using IBM SPSS Statistics: An Interactive Hands-On Approach Second Edition. Sage Publications.
- Alkin, M., Alkin, M., & Christie, C. (2012). An Evaluation Theory Tree. Evaluation Roots, May, 13-65.
- Allen, C. S., Chen, Q., Willson, V. L., & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multi-level analysis. *Educational Evaluation and Policy Analysis,* 31(4), 480-499. <u>https://doi.org/10.3102/0162373709352239</u>
- Almeida, M. A. P. de. (2014). As epidemias nas notícias em Portugal: cólera, peste, tifo, gripe e varíola, 1854-1918. *História, Ciências, Saúde – Manguinhos, 21*(2), 687–708.
- Aloisi, C. (2023). The future of standardised assessment: Validity and trust in algorithms for assessment and scoring. *European Journal of Education*, *58*(1), 98-110.
- Alves, I., Campos Pinto, P., & Pinto, T. J. (2020). Developing inclusive education in Portugal: Evidence and challenges. *Prospects*, *49*(3–4), 281–296.
- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Babbie, E. (2013). The Practice of Social Research (13th ed.). Cengage Learning.
- Baird, J. A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: fields apart? Assessment in Education: Principles, Policy & Practice, 24(3), 317–350. <u>https://doi.org/10.1080/0969594X.2017.1319337</u>
- Baker, E. L. (2001) Testing and assessment: A progress report. Educational Assessment, 7(1), 1-12,
- Baker, E.L. (2003) Reflections on Technology-enhanced Assessment. Assessment in Education: Principles, Policy & Practice, 10:3, 421-424
- Ball, S. J. (2008). The education debate. Policy press.
- Bardin, L. (2014). Análise de conteúdo. Edições 70.
- Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. In H. Fives & M. Gregoire Gill (Eds.), *International Handbook of Research on Teacher Beliefs* (pp. 284-300). Routledge.
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, O., & Mariman, R. (2024). Generative Al can harm learning.
- Benavente, A., Peixoto, P., & Gomes, R. M. (2020). *Impacto da covid-19 no sistema de ensino português: Resultados parciais a 12 de abril de 2020*. Observatório de Políticas de Educação e Formação.

- Bennett, R. E. (2011). Formative assessment: a critical review. Assessment in Education: Principles, Policy & Practice, 18: 1, 5 25 http://dx.doi.org/10.1080/0969594X.2010.513678
- Bennett, R.E. (2001). How the Internet will help large-scale assessment reinvent itself. *Education Policy Analysis Archives*, 9(5).
- Berman, A. I., Haertel, E. H., & Pellegrino, J. W. (2020). Introduction: Framing the issues. In Berman, A. I., Haertel,
 E. H., & Pellegrino, J. W. (Eds.), *Comparability of large-scale educational assessments: Issues and recommendations* (pp. 9–24). National Academy of Education. <u>https://doi.org/10.31094/2020/1</u>
- Black, P., & Wiliam, D. (1998a). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy* & *Practice*, 5(1), 7–74.
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappa International*, *80*(2), 139–144. <u>http://www.jstor.org/stable/20439383</u>
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy* & *Practice*, 1–25.
- Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. *Teachers College Record*, *70*(10), 26-50.
- Boin, A., & Hart, P. T. (2003). Public leadership in times of crisis: mission impossible?. *Public administration review*, *63*(5), 544-553.
- Broadfoot, P. M., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002). *Assessment for learning: 10 principles.* Cambridge, UK: University of Cambridge School of Education
- Brookhart, S. M. (2013). Grading. In J. H. McMillan (Ed.), *Sage handbook of Research on Classroom Assessment* (pp. 257–271). Sage Publications, Inc.
- Brown, G. T. L. (2022). The past, present and future of educational assessment: A transdisciplinary perspective. *Frontiers in Education.*
- Bryman, A. (2012). Social research methods (4th ed.). Oxford University Press.
- Buzick, H. M. (2019). Testing Accommodations and the Measurement of Student Academic Growth, *Educational Assessment*, 24:1, 57-72. DOI: <u>https://doi.org/10.1080/10627197.2018.1545571</u>
- Caria, T. H. (2000). A cultura profissional dos professores: o uso do conhecimento em contexto de trabalho na conjuntura da reforma educativa dos anos 90 [Teachers' professional culture: the use of knowledge in the work context in the 90s educational reform]. Fundação Calouste Gulbenkian.
- Carmo, R. M.; Tavares, I.; Cândido A. F. (Ed.) (2020). *Um Olhar Sociológico sobre a Crise Covid-19 em Livro* [A Sociological Glance at the Covid-19 Crisis in a Book]. Observatório das Desigualdades, CIES-Iscte.
 - https://www.observatorio-das-desigualdades.com/2020/11/29/umolharsociologicosobreacovid19emlivro/
- Cavaco, C., Alves, N., Guimarães, P., Feliciano, P., Paulos, C. (2021) Teachers' perceptions of school failure and dropout from a gender perspective: (re)production of stereotypes in school. *Educ Res Policy Prac* 20, 29–44. https://doi.org/10.1007/s10671-020-09265-7
- Cipriano, G., & Martins, S. C. (2021). Beliefs on assessment and grade repetition among teachers in Portugal. *Meta: Avaliação*, 13(39), 248–273. <u>http://dx.doi.org/10.22347/2175-2753v13i39.3033</u>

- Cipriano, G., & Martins, S. C. (2023). Avaliação das aprendizagens em Portugal durante a pandemia. In D. Fonseca,
 A. Neto-Mendes, M. Gonçalves, A. Ventura, & J. A. Costa (Orgs.), A gestão da incerteza na educação:
 (Des)continuidades e desafios: Livro de atas do X Simpósio de Organização e Gestão Escolar (pp. 191-210).
 UA Editora.
- Cipriano, G., & Martins, S. C. da (2024). Introducing technologies into national large-scale testing: Are we ready? *Education Policy Analysis Archives*, 32(20). <u>https://doi.org/10.14507/epaa.32.8538</u>
- Clark, L. A., & Watson, D. (1995). Constructing Validity: Basic Issues in Objective Scale Development. *Psychological Assessment*, 7(3), 309–319.
- Clarke, A. (1999). Evaluation research: An introduction to principles, methods and practice. Sage.
- CNE. (2021a). Educação em tempo de pandemia: Problemas, respostas e desafios das escolas. Conselho Nacional de Educação.
- CNE. (2021b). *Efeitos da pandemia COVID-19 na educação: Desigualdades e medidas de equidade.* Conselho Nacional de Educação.
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, 112(1), 155–159. <u>https://doi.org/10.1037/0033-</u> 2909.112.1.155
- Coladarci, T. (2002). Is it a house...or a pile of bricks? Important features of a local assessment System. *Phi Delta Kappan 83*(10), 772–774.
- Cooper, A., Deluca, C., Holden, M., & Macgregor, S. (2022). Emergency assessment: rethinking classroom practices and priorities amid remote teaching. *Assessment in Education: Principles, Policy & Practice*
- Costa, A.F., Pegado, E., Ávila, P., Coelho, A.R. (2015). Evaluating the Portuguese National Reading Plan: teachers' perceptions on the impact in schools. *Educ Res Policy Prac* 14, 119–138. <u>https://doi.org/10.1007/s10671-014-9171-y</u>
- Costa, E., Baptista, M. (2023). Reframing Schools: What Has Been Learned and Remains in the Post-COVID-19 Period in Portugal. In Reimers, F. M. (Ed.), *Schools and Society During the COVID-19 Pandemic: How Education Systems Changed and the Road Ahead* (pp. 149-167). Springer.
- Council of Ministers Announcement (2024). *Avaliar melhor, aprender mais: Novo modelo de avaliação externa 2024/2025 e anos seguintes* [Assess better, learn more: New external assessment model for 2024/2025 and beyond]. *Ministério da Educação, Ciência e Inovação*.
- Crahay, M., Wanlin, P., Issaieva, É., & Laduron, I. (2010). Fonctions, structuration et évolution des croyances (et connaissances) des enseignants [Functions, structuring and evolution of teachers' beliefs (and knowledge)]. *Revue française de pédagogie. Recherches en éducation*, (172), 85-129. <u>https://doi.org/10.4000/rfp.2296</u>
- Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- Dermo, J. (2009). e-Assessment and the student learning experience: A survey of student perceptions of eassessment. *British Journal of Educational Technology*, *40*(2), 203-214.
- DGAE (2019). Códigos de Agrupamentos de Escolas / Escolas não Agrupadas da Rede do Ministério da Educação. Direção-Geral da Administração escolar

- DGE (2023). *Orientações Curriculares para a Educação Pré-Escolar* [Curricular Guidelines for Early Childhood Education]. Direção Geral de Educação.
- Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory into practice*, 55(2), 153-159. https://doi.org/10.1080/00405841.2016.1148989
- DRAEM (2023). *Quadros sínteses com estatísticas da educação.* Direção Regional de Administração Escolar da Madeira.
- DREAEA (2023). *Estatísticas da Educação*. Direção Regional de Educação e Administração Educativa dos Açores.
- Eagly, A. H., & Chaiken, S. (1993). The Psychology of Attitudes. Harcourt Brace Jovanovich College Publishers.
- Engzell, P., Frey, A., & Verhagen, M. (2020). Learning inequality during the COVID-19 pandemic. 1–45.
- Ercikan, K., Asil, M., & Grover, R. (2018). Digital divide: A critical context for digitally based assessments. *Education Policy Analysis Archives*, *26*(51). <u>https://doi.org/10.14507/epaa.26.3817</u>
- Esteves, M., Freitas, P., Herdade, M., Carvalho, B. P., & Peralta, S. (2021). *Crianças em Portugal e ensino a distância: um retrato.* Zenodo.
- Fernandes, D. (2009) Educational assessment in Portugal, Assessment in Education: Principles, Policy & Practice, 16:2, 227-247
- Fernandes, D. (2010). Acerca da Articulação de Perspetivas e da Construção Teórica em Avaliação Educacional.
 In M. T. Esteban & J. A. Almerindo (Eds.), *Olhares E Interfaces: Reflexões Críticas Sobre A Avaliação* (pp. 15–44). Cortez Editora.
- Fernandes, D. (2011). Articulação da aprendizagem, da avaliação e do ensino: Questões teóricas, práticas e metodológicas. In M. P. A. e J.-M. D. K. (Orgs.) (Ed.), *Do currículo à avaliação, da avaliação ao currículo* (pp. 131–142). Porto Editora.
- Fernandes, D. (2013). Avaliação em educação: Uma discussão de algumas questões críticas e desafios a enfrentar nos próximos anos. *Ensaio*, *21*(78), 11–34.
- Fernandes, D. (2014a). Avaliação das aprendizagens e políticas educativas: O difícil percurso da inclusão e da melhoria. In M. L. Rodrigues (Ed.), *40 Anos de Políticas de Educação em Portugal: volume I, A construção do sistema democrático de ensino* (pp. 231–268). Edições Almedina.
- Fernandes, D. (2014b). Avaliações externas e melhoria das aprendizagens dos alunos: Questões críticas de uma relação (im)possível. In Ferreira, A. (Ed.), Avaliação Externa e Qualidade das Aprendizagens, 21-49.
 Conselho Nacional de Educação
- Fernandes, D. (2019a). Avaliações externas e aprendizagens dos alunos: uma reflexão crítica. Linhas Críticas.
- Fernandes, D. (2019b). Contributos das perspetivas orientadas por / para uma agenda social. *Metodologias de Avaliação de Políticas Públicas, January 2018,* 49–67.
- Fernandes, D. (2021). Formação contínua de professores em tempos pandémicos: o caso do Projeto MAIA. *Linhas Críticas*, 27. <u>https://doi.org/10.26512/lc27202139025</u>
- Fernandes, D., Machado, E., & Candeias, F. (2021). *Para uma Avaliação Pedagógica: Dinâmicas e Processos de Formação no Projeto MAIA (2020-2021).* Ministério da Educação/Direção-Geral da Educação.
- Field, A. (2013). Discovering statistics using IBM SPSS statistics. Sage.

- Flores, M. A., & Day, C. (2006). Contexts which shape and reshape new teachers' identities: A multi-perspective study. *Teaching and teacher education*, *22*(2), 219-232.
- Flores, M. A., & Gago, M. (2020). Teacher education in times of COVID-19 pandemic in Portugal: national, institutional and pedagogical responses. *Journal of Education for Teaching*.
- Flores, M. A., Barros, A., Gago, M., Fernandes, E. L., Pereira, D., Ferreira, P. C., & Costa, L. (2021). Emergency remote teaching in pandemic times: The experience of Portuguese teachers. *Revista Portuguesa de Investigação Educacional*, (21), 1-26. <u>https://doi.org/10.34632/investigacaoeducacional.2021.10022</u>
- Flores, M. A.; Barros A; Simão, A. M. V. S.; Pereira D.; Gago M.; Fernandes, E. L.; Costa Ferreira, P. C. & Costa, L. (2023): Remote teaching in times of COVID-19: teachers' adaptation and pupil level of participation. *Technology, Pedagogy and Education*. <u>https://doi.org/10.1080/1475939X.2023.2270491</u>
- Formosinho, J., & Machado, J. (2013). A regulação da educação em Portugal do Estado Novo à democracia [The regulation of education in Portugal—from the Estado Novo to democracy]. *Educação: Temas e Problemas*, (12 13), 27-40.
- Gamlem, S. M., Smith, K., & Sandvik, L. V. (2023). Stakeholder Opinions About Cancelling Exams in Norwegian Upper Secondary School During the Pandemic, and its Consequences–An Illuminative Study. *Assessment Matters*.
- Gardner, J., Harlen, W., Hayward, L., Stobart, G., & Montgomery, M. (2010). *Developing teacher assessment*. McGraw-Hill Education
- Geijsel, F., Sleegers, P., Van Den Berg, R., & Kelchtermans, G. (2001). Conditions Fostering the Implementation of Large-Scale Innovation Programs in Schools: Teachers' Perspectives. *Educational Administration Quarterly*, 37(1), 130–166.
- Gibson, J. P. Krimmer, R., Teague, V., & Pomares, J. (2016). A review of e-voting: the past, present and future. Annals of Telecommunications, 71, 279-286.
- Gipps, C. (2011). Beyond Testing (Classic Edition): Towards a theory of educational assessment. Routledge.
- Gipps, C., Stobart, G. (2003). Alternative assessment. In T. Kellaghan e D. Stufflebeam (Eds), *International handbook of educational evaluation*, 549-575. Dordrecht: Kluwer.
- Goldring, L. (2002). The power of school culture. *Leadership*, *32*(2), 32-35.
- Greiff, S., Wüstenberg, S., Holt, D. V., Goldhammer, F., & Funke, J. (2013). Computer-based assessment of complex problem solving: Concept, implementation, and application. *Educational Technology Research and Development*, *61*, 407-421.
- Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. John Wiley & Sons.
- Guba, E. G., & Lincoln, Y. S. (1989). Fourth Generation Evaluation. Sage Publications.
- Hair, J. F., Black, W. C., Badin, B. J., Anderson, R. E., & Tatham, L. R. (2006). *Multivariate Data Analysis*. Pearson Education.
- Hair, J., Anderson R., e Black, W. (2019). *Multivariate Data Analysis: A Global Perspective* (8th edition). Pearson International Edition

- Harrell, F. E. (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis.* Springer.
- Hayes, A. F. (2022). Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach (3rd edition). Guilford Press.
- Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2024, May). *The difference between emergency remote teaching and online learning.* EDUCAUSE Review.
- Hopfenbeck, T. N. (2018). Classroom assessment, pedagogy and learning twenty years after Black and Wiliam 1998. *Assessment in Education: Principles, Policy & Practice, 25*(6), 545–550.

https://doi.org/10.1080/0969594X.2018.1553695

- IAVE. (2020). Provas de Avaliação Externa | 2019/2020. Instituto de Avaliação Educativa, I.P., 1–14.
- IAVE. (2021). Estudo Diagnóstico das Aprendizagens. Ministério Da Educação, 1–22.
- IAVE. (2022). O DAVE Desmaterialização da Avaliação Externa. Instituto de Avaliação Educativa, 1-2.
- Ikeda, M., & Echazarra, A. (2021). How socio-economics plays into students learning on their own: Clues to COVID-19 learning losses. In *PISA in Focus* (Vol. 114). OECD Publishing.
- Jensen, R. E., Roe, A., & Blikstad-Balas, M. (2024). The smell of paper or the shine of a screen? Students' reading comprehension, text processing, and attitudes when reading on paper and screen. *Computers & Education*, *219*, 105107.
- Jerrim, J., Allen, R., & Sims, S. (2024). How did the COVID-19 pandemic affect the anxiety of teachers at work? *Educational Review*. Advance online publication. <u>https://doi.org/10.1080/00131911.2023.2293455</u>
- Jimerson, S. R. (2001a). A Synthesis of Grade Retention Research: Looking Backward and Moving Forward. *Contemp School Psychol* 6, 47–59. <u>https://doi.org/10.1007/BF03340883</u>
- Jimerson, S. R. (2001b). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School psychology review*, *30*(3), 420-437. https://doi.org/10.1080/02796015.2001.12086124
- Johnson, M.; Shaw, S. (2018) What is computer-based testing washback, how can it be evaluated and how can this support practitioner research? *Journal of Further and Higher Education*
- Justino, D. (2014). Escolaridade obrigatória: entre a construção retórica e a concretização política [Compulsory Education: Between Rhetorical Construction and Political Realization] In Rodrigues M.L. (Ed.) 40 anos de políticas de educação em Portugal – Volume I: A construção do sistema democrático de ensino [40 years of education policies in Portugal – Volume I: The construction of the democratic education system]. Almedina.
- Kellagham, T.; Madaus, G. (2003). External (public) examinations. In T. Kellaghan e D. Stufflebeam (Eds), International handbook of educational evaluation (9th ed), pp. 577-602. Kluwer Academic Publishers.
- Kellaghan, T., Stufflebeam, D. L., & Wingate, L. A. (2003). Introduction. In T. Kellaghan, D. L. Stufflebeam, & L. A.
 Wingate (Eds.), *International Handbook of educational Evaluation* (9th ed), pp. 1–6. Kluwer Academic Publishers.
- Kerrey, B., & Isakson, J. (2000). *The power of the Internet for learning: Moving from promise to practice.* (Report of the Web-based Education Commission). Washington, D.C.: Web-Based Education Commission.
- Kline, R. B. (2016). Principles and practice of structural equation modelling (4th edition). Guilford publications.
- Klusmann, U., Aldrup, K., Roloff-Bruchmann, J., Carstensen, B., Wartenberg, G., Hansen, J., & Hanewinkel, R.
 (2023). Teachers' emotional exhaustion during the COVID-19 pandemic: Levels, changes, and relations to pandemic-specific demands. *Teaching and Teacher Education*, 121.
- Kothari, C. R. (2004). Research methodology: Methods and techniques. New Age International Publishers.
- Lambe, P. (2007). Organising knowledge: taxonomies, knowledge and organisational effectiveness. Chandos Publishing.
- Le Fevre, D., & Richardson, V. (2002). Staff development in early reading intervention programs: The facilitator. *Teaching and Teacher Education*, *18*(4), 483-500. <u>https://doi.org/10.1016/S0742-051X(02)00011-2</u>
- Lee, J.; Lee, J. S.; Ellis, N. (2023). Public Opinion About National Large-Scale Student Assessment: A Case of NAPLAN, *Educational Assessment*, 28:3, 137-155. https://doi.org/10.1080/10627197.2023.2215977

Lingard, B. (2007). Pedagogies of indifference. International Journal of Inclusive Education, 11(3), 245–266.

- Linn, R. L. (2001). A Century of Standardized Testing: Controversies and Pendulum Swings, *Educational* Assessment, 7:1, 29-38. <u>https://doi.org/10.1207/S15326977EA0701_4</u>
- Lizana, P. A., Vega-Fernadez, G., Rodero-Cosano, M. L., Salinas-Perez, J. A., Conejo-Cerón, S., Bagheri, N., & Castelpietra, G. (2021). Teacher Teleworking during the COVID-19 Pandemic: Association between Work Hours, Work-Family Balance and Quality of Life. *International Journal of Environmental Research and Public Health*, 18, 7566
- Lomba, P. (2020). Constitution, state of emergency and public health Administration: some problems. *e-Pública*, *7*(1), 27-43.
- Lourtie, P. (2020). *Recomendação: O acesso ao ensino superior e a articulação com o ensino secundário*. Conselho Nacional de Educação.
- Machado, E. A., Flores, M. A., Pereira, D., Fernandes, E., & Costa, L. (2022). Políticas e práticas de avaliação externa dos alunos na perspetiva dos diretores: o caso das provas de aferição. *Revista Portuguesa de Investigação Educacional*, 23, 1–26.
- Madaus, G. (1993). A National Testing System: Manna From Above? An Historical/Technological Perspective, *Educational Assessment*, 1:1, 9-26. <u>https://doi.org/10.1207/s15326977ea0101_2</u>
- Madaus, G., & Russell, M. (2010). Paradoxes of High-Stakes Testing. Journal of Education, 190, 21–30.
- MAIA Project (2023). National Project of Training, Supervision and Research in Classroom Assessment (MAIA). Direção-Geral da Educação. through <u>https://afc.dge.mec.pt/sites/default/files/2023-08/MAIA%20-</u> <u>%20Brochura Inglês.pdf</u>
- Maldonado, J. E., & De Witte, K. (2020). The effect of school closures on standardised student test outcomes. FEB Research Report Department of Economics, September.
- Marôco, J. (2021). *Análise estatística com o SPSS statistics* [Statistical analysis with SPSS statistics] (8th Edition). Report number.
- Martins, S. C. & Albuquerque A. (2019). Autonomy and leadership of school actors in Martins, S. C.; Capucha, L.; Sebastião, J. (Org.), *School autonomy, organization and performance in Europe: A comparative analysis for the period from 2000 to 2015* (pp. 23-32). CIES-IUL.

- Martins, S. C. (2017). Expansão educativa, oportunidades sociais e desempenhos dos sistemas educativos: um olhar cruzado entre os dois lados do Atlântico in Setton, M.G. et al, (Ed.), *Mérito, desigualdades e diferenças: cenários de (in)justiça escolar no Brasil e em Portugal*, Alfenas Minas Gerais, Universidade Federal de Alfenas UNIFAL-MG.
- Martins, S. C. (2020). A educação e a covid-19: Desigualdades, experiências e impactos de uma pandemia não anunciada. In R. M. do Carmo, I. Tavares, & A. F. Cândido (Eds.), *Um olhar sociológico sobre a crise covid-19 em livro* (pp. 37–55). Observatório das Desigualdades, CIES-ISCTE.
- Martins, S. da C.; Mauritti, R.; Nunes, N.; Romão, A. L.; Costa, A. F. (2016). A educação ainda é importante para a mobilidade social? Uma perspetiva das desigualdades educacionais da Europa do Sul no contexto europeu.
 [Is education still important for social mobility? A perspective on educational inequalities in Southern Europe within the European context]. *Revista Portuguesa de Educação*. 29 (2), 261-285.
 https://doi.org/10.21814/rpe.7920
- McLeskey, J., & Waldron, N. L. (2002). Inclusion and School Change: Teacher Perceptions Regarding Curricular and Instructional Adaptations. *Teacher Education and Special Education*, 25(1), 41–54.
- Messick, S. (1996). Validity and washback in language testing. Language testing, 13(3), 241-256.
- Mons, N. (2009). Effets théoriques et réels des politiques d'évaluation standardisée. *Revue française de pédagogie*.
- Muskin, J. (2017). Continuous assessment for improved teaching and learning: A critical review to inform policy and practice Programme and Meeting Document.
- National Academy of Education (NAE). (2021). *Educational Assessment in the covid-19 era and beyond*. National Academy of Education, 1–22.
- National Research Council (NRC). (2001a). *Classroom assessment and the National Science Education Standards*. National Academies Press.
- National Research Council (NRC). (2001b). *Knowing what students know: The science and design of educational assessment.* National Academy Press.
- Nisbet, I., & Shaw, S. (2022). Fair high-stakes assessment in the long shadow of Covid-19. Assessment in Education: Principles, Policy & Practice.
- Nunes, L. C., Reis, A. B., & Seabra, C. (2018): Is retention beneficial to low-achieving students? Evidence from Portugal, *Applied Economics*, https://doi.org/10.1080/00036846.2018.1444261
- OECD. (2020a). Lessons for Education from COVID-19: A Policy Maker's Handbook for More Resilient Systems. OECD Publishing
- OECD. (2020b). School education during covid-19: Were teachers and students ready? Portugal. OECD, 1–9.
- OECD. (2021). The state of school education: One year into the COVID pandemic. OECD Publishing, March.
- OECD. (2022). Mending the Education Divide: Getting Strong Teachers to the Schools That Need Them Most, TALIS, OECD Publishing, Paris.
- OECD. (2023a). Education at a Glance 2023: OECD Indicators, OECD Publishing.
- OECD. (2023b). PISA 2022 Results (Volume II): Learning During and From Disruption, PISA, OECD Publishing.

Ofqual (2020), Online and on-screen assessment in high stakes, sessional qualifications: A review of the barriers to greater adoption and how these might be overcome. Ofqual, London.

Oldfield, A., Broadfoot, P.; Sutherland, R.; Timmins, S. (2012). Assessment in a Digital Age: A research review.

- Oliveira, Maria Marly de (2005). *Como fazer: projetos, relatórios, monografias, dissertações e teses* (3ª ed). Rio de Janeiro: Elsevier.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of educational research*, *62*(3), 307-332. <u>https://doi.org/10.3102/00346543062003307</u>
- Palumbo, R. (2020), "Let me go to the office!" An investigation into the side effects of working from home on work-life balance, *International Journal of Public Sector Management*, Vol. 33 No. 6/7, pp. 771-790.
- Panadero, E., Fraile, J., Pinedo, L., Rodríguez-Hernández, C., & Díez, F. (2022). Changes in classroom assessment practices during emergency remote teaching due to COVID-19. *Assessment in Education: Principles, Policy & Practice, 29*(3), 361–382.
- Parveva, T., Horváth, A., Krémó, A., & Sigalas, E. (2020) *Eurydice Brief. Equity in school education in Europe: structures, policies and student Performance.*, European Commission.
- PASEO (2017). *Perfil dos alunos à saída da escolaridade obrigatória*. Ministério da Educação / Direção-Geral da Educação
- Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and teacher education*, *84*, 128-138. <u>https://doi.org/10.1016/j.tate.2019.05.003</u>
- Patil, Vivek H., Surendra N. Singh, Sanjay Mishra, & D. Todd Donavan (2007). Parallel Analysis Engine to Aid Determining Number of Factors to Retain [Computer software].
- Pedrosa, J., Leite, C., da Rosa, M. J. P., Pinho, I., Vieira, C., & Marinho, P. (2022). *Avaliação de aprendizagens em instituições educativas* [Learning Assessment in Educational Institutions]. Fundação Calouste Gulbenkian
- Pittas, E., & Papanastasiou, E. (2023). Effects of COVID-19 on the educational performance of children with special educational needs and disabilities: A systematic review according to children's/youth's and caregivers' perspectives. *Research in Developmental Disabilities*, 143, 104635. https://doi.org/10.1016/j.ridd.2023.104635
- Popping, R. (2015). Analyzing open-ended questions by means of text analysis procedures. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *128*(1), 23-39.
- Rädiker, S., & Kuckartz, U. (2020). Focused Analysis of Qualitative Interviews With MAXQDA: Step by Step. Maxqda Press.
- Reis, A. B., Lima, G., Nunes, L. C., Freitas, P., & Alves, T. (2020). *Ensino a Distância: 2º Questionário a Professores,* 1–14. Nova School of Business & Economics.
- Ribeiro, V. M., Kasmirski, P. R., Gusmão, J. B. de, Batista, A. A. G., Janomini, M. A., & Crahay, M. (2018). Crenças de professores sobre reprovaça~o escolar. *Educação Em Revista*, 34, Article 173086. https://doi.org/10.1590/0102-4698173086
- Richardson, M., & Clesham, R. (2021). Rise of the machines? The evolving role of AI technologies in high-stakes assessment. *London Review of Education*, 19(1), 0–13
- Richardson, V., & Placier, P. (2001). Teacher change. Handbook of research on teaching, 4, 905-947.

- Rodrigues, C., Costa, J. M., & Moro, S. (2022). Assessment Patterns during Portuguese Emergency Remote Teaching. *Sustainability*, 14(5), 3131.
- Rodrigues, M. L. (Ed.) (2014a). 40 anos de políticas de educação em Portugal Volume I: A construção do sistema democrático de ensino [40 years of education policies in Portugal – Volume I: The construction of the democratic education system]. Almedina.
- Rodrigues, M. L. (Ed.) (2014b). 40 Anos De Políticas De Educação Em Portugal Volume II: Conhecimento, Atores E Recursos [40 years of education policies in Portugal - Volume II - Knowledge, actors and resources] Almedina.
- Rodrigues, M. L., J. Sebastião, J. T. Mata, L. Capucha, L. Araújo, M Vieira da Silva, Susana da Cruz Martins, & Valter Lemos (2016). *Educação. 30 Anos de Lei de Bases* [Education. 30 Years of the Education Act]. Mundos Sociais.
- Rogne, W. M., Rønneberg, V., Gamlem, S. M., Spilling, E. F., & Uppstad, P. H. (2024). Effects of digitalisation on learning to write–A naturalistic experiment. *Learning and Instruction*, *93*, 101970.
- Rozenwajn, E. & Dumay, X. (2014). Les effets de l'évaluation externe sur les pratiques enseignantes: une revue de la literature. *Revue française de pédagogie.*
- Sá, C., Sin, C., Pereira, F., Aguiar, J., & Tavares, O. (2021). Estudantes nacionais e internacionais no acesso ao ensino superior: Quem são, que escolhas fazem e como acedem ao mercado de trabalho. EDULOG Fundação Belmiro de Azevedo.
- Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in education: principles, policy & practice, 5*(1), 77-84. <u>http://dx.doi.org/10.1080/0969595980050104</u>
- Salmons, J. E. (2022). Doing Qualitative Research Online (Second Edition). Sage Publications Ltd.
- Sandvik, L. V., Svendsen, B., Strømme, A., Smith, K., Sommervold, O. A., & Angvik, S. A. (2023). Assessment during COVID-19: Students and Teachers in Limbo When the Classroom Disappeared. *Educational Assessment*, 28(1), 11-26. <u>https://doi.org/10.1080/10627197.2022.2122953</u>
- Santana, M. (2019). Práticas e representações acerca da retenção escolar [Practices and representations regarding grade retention] (Doctoral dissertation). *Lisboa: Faculdade de Ciências Sociais e Humanas*. <u>https://run.unl.pt/bitstream/10362/89715/1/Tese Práticas%20e%20Representações%20acerca%20da%2</u> <u>ORetenção%20Escolar.pdf</u>
- Santiago, P., Donaldson, G., Looney, A. e Nusche, D. (2012). *OECD Reviews of evaluation and assessment in education: Portugal 2012.* Paris: OECD Publishing.
- Santos, N. N., & Monteiro, V. (2023). Grade retention: teachers' beliefs and practices and their relationship to school educational policies. *Ensaio: Avaliação e Políticas Públicas em Educação*, *31*(121). https://doi.org/10.1590/S0104-40362023003103925
- Santos, N. N., Pipa, J., & Vera Monteiro. (2023). Analysing grade retention beliefs within teachers' psychopedagogic beliefs system. *Teaching and Teacher Education*, 121, 103939. https://doi.org/10.1016/j.tate.2022.103939
- Schuck, R.K.; Lambert, R. (2020). "Am I Doing Enough?" Special Educators' Experiences with Emergency Remote Teaching in Spring 2020. *Educ. Sci.* 2020, *10*, 320.
- Scriven, M. (1994). Evaluation as a discipline. Studies in Educational Evaluation, 20(1), 147–166.

- Seabra, F.; Teixeira, A.; Abelha, M.; Aires, L. (2021) Emergency Remote Teaching and Learning in Portugal: Preschool to Secondary School Teachers' Perceptions. *Educ. Sci.* 2021, 11, 349.
- Seidman, I. (2006). Interviewing as qualitative research: A guide for researchers in education and the social sciences. Teachers College Press.
- SGEC (2022). Orientação técnica investimento C20- i01 "transição digital na educação". Secretaria-Geral da Educação e Ciência.
- Shaw, S., & Crisp, V. (2015). Reflections on a framework for validation-Five years on. Cambridge assessment.
- Simó-Pinatella, D., Goei, S. L., Carvalho, M., & Nelen, M. (2022). Special education teachers' experiences of addressing challenging behaviour during the pandemic. *European Journal of Special Needs Education*, *36*(6), 907–920.
- Slomp, D., Marynowski, R., Holec, V. et al. (2020) Consequences and outcomes of policies governing mediumstakes large-scale exit exams. *Educ Asse Eval Acc* 32, 431–460.
- Smith, W.C., Holloway, J. (2020). School testing culture and teacher satisfaction. Educ Asse Eval Acc 32, 461–479.
- Souza, C. (2006). Políticas públicas: uma revisão da literatura. Sociologias, 16, 20–45.
- Stake, R., & Schwandt, T. (2006). On discerning quality in evalution. In I. Shaw, J. E. Greene, & M. Mark (Eds.), *The Sage handbook of evaluation* (pp. 404–418). Sage.
- Stobart, G., & Eggen, T. (2012). High-stakes testing value, fairness and consequences. *Assessment in Education: Principles, Policy and Practice*, 19(1), 1–6.
- Stopher, P. (2012). Collecting, managing, and assessing data using sample surveys. In *Collecting, Managing, and Assessing Data Using Sample Surveys*. Cambridge University Press.
- Strauss, A. & Corbin, J. (2015). Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Sage publications.
- Stufflebeam, D. L. (1999). Foundational Models for 21 st Century Program Evaluation by Foundational Models for 21 st Century Program Evaluation. *Canadian Journal of Family and Youth*, 7(1), 27–54.
- Stufflebeam, D. L. (2003). The CIPP model for evaluation. In T. Kellaghan e D. Stufflebeam (Eds), *International handbook of educational evaluation*, 31-62. Dordrecht: Kluwer.
- Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation, Theory, Models & Applications* (second edition). Jossey-Bass.
- Stufflebeam, D. L., & Shinkfield, A. J. (1985). *Systematic evaluation: A self-instructional guide to theory and practice*. Kluwe-Nijhoff Publishing.
- Sudman, S., & Bradburn, N. M. (1982). Asking questions: A practical guide to questionnaire design. Jossey-Bass.
- Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., ... & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, *3*, 100075.
- Tabachnick, B. G., & Fidell, L. S. (2018). Using multivariate statistics (7th ed.). Pearson.
- Taras, M. (2005). Assessment summative and formative some theoretical reflections. *British journal of educational studies*, *53*(4), 466-478. <u>https://doi.org/10.1111/j.1467-8527.2005.00307.x</u>
- Taras, M. (2012). Assessing assessment theories. Online Educational Research Journal, 3(12).

- Tomas, C., Borg, M., & McNeil, J. (2015). E-assessment: Institutional development strategies and the assessment life cycle. *British Journal of Educational Technology*, *46*(3), 588-596.
- Tomasik, M. J., Helbling, L. A., & Moser, U. (2020). Educational Gains of In-Person vs. Distance Learning in Primary and Secondary Schools: A Natural Experiment During the COVID-19 Pandemic School Closures in Switzerland. *International Journal of Psychology*.
- Tomchin, E. M., & Impara, J. C. (1992). Unraveling teachers' beliefs about grade retention. *American Educational Research Journal*, *29*(1), 199-223. <u>https://doi.org/10.3102/00028312029001199</u>
- Torres, L. L., & Alves, M. G. (2024). Democratisation and educational inclusion during lockdown times: Perceptions of Portuguese teachers. *Education Sciences*, *14*(2), 197. <u>https://doi.org/10.3390/educsci14020197</u>
- Torres, L. L., Palhares, J. A., & Afonso, A. J. (2019). The distinction of excellent students in the Portuguese state school as a strategy of educational marketing accountability. *Educational assessment, evaluation and accountability*, *31*, 155-175.
- Tsagari, D., Cheng, L. (2017). Washback, Impact, and Consequences Revisited. In: Shohamy, E., Or, I., May, S. (eds) Language Testing and Assessment. Encyclopedia of Language and Education. Springer, Cham.

UNESCO (2009). Policy Guidelines on Inclusion in Education. Paris, UNESCO.

Vartanian, T. P. (2011). Secondary Data Analysis. Oxford University Press.

- Verdasca, J. (2013), Rankings escolares: «a César o que é de César e a Deus o que é de Deus». *Educação: Temas e Problemas*, 12-13, 294-327.
- Wiese, E., & Nortvedt, G. A. (2023). Teacher assessment literacy in culturally and linguistically diverse classrooms:
 A Norwegian case study. *Teaching and Teacher Education*, *135*, 104357.
 https://doi.org/10.1016/j.tate.2023.104357
- Wiese, E., & Nortvedt, G. A. (2023). Teacher assessment literacy in culturally and linguistically diverse classrooms:
 A Norwegian case study. *Teaching and Teacher Education*, *135*, 104357.
 https://doi.org/10.1016/j.tate.2023.104357
- Wiliam, D. (2006). Formative Assessment: Getting the Focus Right. *Educational Assessment* 11(3): 283-289. http://dx.doi.org/10.1080/10627197.2006.9652993
- Wills, G., & van der Berg, S. (2024). COVID-19 disruptions and education in South Africa: Two years of evidence. Development Southern Africa, 41(2), 446-465. <u>https://doi.org/10.1080/0376835X.2024.2311711</u>
- Womack, T. A., & Monteiro, E. M. (2022). Special education staff well-being and the effectiveness of remote services during the COVID-19 pandemic. *Psychology in the Schools*.
- Xu, Y. (2013). Classroom Assessment in Special Education. In J. H. McMillan (Ed.), *Sage Handbook of Research on Classroom Assessment* (pp. 431–447). Sage Publications, Inc.

Other Sources

- Antena Livre through <u>https://www.antenalivre.pt/covid-19/cronologia-covid-19-principais-acontecimento-da-</u> pandemia-em-portugal
- CNE Conselho Nacional de Educação through https://www.cnedu.pt/en/
- Correio da Manhã through <u>https://www.cmjornal.pt/sociedade/detalhe/veja-a-cronologia-dos-principais-</u> acontecimentos-desde-o-inicio-da-pandemia-de-coronavirus
- CRUP Conselho de Reitores das Universidades Portuguesas through https://www.crup.pt
- DGAE Direção-Geral da Administração Escolar through https://www.dgae.medu.pt/download/recrutamento/
 - outros/201819/20190307 Rec Codigos-AE ENA-2019 2020.pdf
- DGE Direção-Geral da Educação through www.dge.mec.pt
- DGEEC Direção-Geral de Estatísticas da Educação e Ciência through www.dgeec.mec.pt
- DGES Direção-Geral do Ensino Superior through www.dges.gov.pt
- DGS https://covid19.min-saude.pt/numero-de-novos-casos-e-obitos-por-dia/
- Diário de Notícias through <u>https://www.dn.pt/vida-e-futuro/cronologia-de-uma-pandemia-em-portugues-os-</u> <u>tres-meses-que-mudaram-o-pais-12259916.html/</u>
- EU European Commission through <u>https://commission.europa.eu/business-economy-euro/economic-</u> recovery/recovery-and-resilience-facility/portugals-recovery-and-resilience-plan_en
- Eurydice European Commission through <u>https://eurydice.eacea.ec.europa.eu/national-education-systems/</u> portugal/national-reforms-school-education
- GestEdu Instituto de Gestão Financeira da Educação through https://www.gesedu.pt/PesquisaRede
- IGEC Inspeção Geral da Educação e Ciência through https://www.igec.mec.pt/
- INE Instituto Nacional de Estatística through https://www.ine.pt
- Jornal de Notícias through <u>https://www.jn.pt/nacional/cronologia-dos-principais-acontecimentos-de-um-ano-</u> <u>de-covid-em-portugal-13400044.html/</u>
- PRR *Plano de Recuperação e Resiliência* through <u>https://recuperarportugal.gov.pt/2021/06/13/investimento-</u> td-c20-i01/
- WHO World Health Organization through <u>https://www.who.int/emergencies/diseases/novel-coronavirus-</u> 2019/interactive-timeline#event-52

This page intentionally left blank

Legislation

Council of Ministers Resolution No. 30/2020, of 21 April Decree of the President of the Republic No. 14-A/2020 of 18 March Decree of the President of the Republic No. 17-A/2020 of 02 April Decree of the President of the Republic No. 20-A/2020 of 06 May Decree-Law No. 139/2012 of 05 July Decree-Law No. 14-G/2020 of 13 April Decree-Law No. 15/2007 of 19 January Decree-Law No. 17/2016 of 04 April Decree-Law No. 209/2002 of 17 October Decree-Law No. 22-D/2021 of 22 March Decree-Law No. 22/2023 of 03 April Decree-Law No. 27-B/2022 of 23 March Decree-Law No. 296-A/98, of 25 September Decree-Law No. 54/2018 of 06 July Decree-Law No. 55/2018 of 06 July Decree-Law No. 6/2001, of 18 January Decree-Law No. 62/2023 of 25 July Decree-Law No. 75/2008 of 22 April Decree-Law No. 85/2009 of 27 August Dispatch No. 6478/2017 of 26 July Law No. 46/86 of 14 October Law No. 85/2009 27 August Normative Dispatch No. 1.F/2016 of 05 April Normative Dispatch No. 4/2024 of 21 February Normative Dispatch No. 5437/2000 of 09 March Normative Dispatch No. 644-A/94 of 15 September Normative Dispatch No. 98A/92 of 20 June Order No. 2836-A/2020 of 19 March Order No. 3298-B/2020 of 13 March Order No. 3301-B/2020 of 15 March Ordinance No. 212/2009 of 23 February Regulatory Decree No. 26/2012 of 21 February

This page intentionally left blank

APPENDIXES

This page intentionally left blank

APPENDIX A

Head teachers interview script

ANTES DE COMEÇAR

- 1. Agradecer a disponibilidade e voltar a enquadrar o estudo.
- Garantir que consentimento informado foi assinado, solicitar autorização para a gravação do áudio da entrevista, reiterando a salvaguarda do anonimato do(a) diretor(a) e da escola.
- 3. Permitir a desistência a qualquer momento.

BLOCO 1: AVALIAÇÃO INTERNA EM PANDEMIA

Q1. Quando deflagrou a pandemia covid-19 e as escolas fecharam em março de 2020, as escolas estavam praticamente com as avaliações finalizadas. Que dificuldades sentiram ao nível da avaliação na conclusão desse ano letivo?

Q2. Que medidas foram tomadas ao nível da <u>avaliação sumativa</u> das aprendizagens? DESBLOQUEADORES (se necessário):

- Q2.1 Houve a necessidade de fazer alteração de critérios de avaliação?
- Q2.2 Houve a definição, ao nível de escola, de novos instrumentos de avaliação?
- Q2.3 Houve alguma recomendação para privilegiar a avaliação formativa?
- Q3. Que recomendações foram feitas relativamente à <u>classificação final</u> de ano letivo? DESBLOQUEADORES (se necessário):

Q3.1 Ao nível da classificação final de ano letivo, houve alguma recomendação para privilegiar o domínio comportamental (participação, assiduidade e cumprimento de tarefas) e menos o domínio cognitivo?

Q3.2 Ao nível da classificação final de ano letivo, houve alguma preocupação especial com os alunos que no 2º período já tinham tido uma classificação insatisfatória?

Q3.3 Ao nível da classificação final de ano letivo, houve alguma preocupação especial com os alunos do ensino secundário que iam ser sujeitos a exame?

Q3.4 Ao nível da classificação final de ano letivo, houve alguma preocupação especial com os alunos mais carenciados e que não tiveram as mesmas oportunidades que os colegas?

Q3.5 Ao nível da classificação final de ano letivo, houve alguma preocupação especial com os alunos mais novos do 1º ciclo?

Q4. Nesse ano letivo, as taxas de retenção baixaram na sua escola? Houve alguma recomendação da direção/conselho pedagógico relativamente à <u>transição/retenção de ano</u>? Se sim, quais?

BLOCO 2: AVALIAÇÃO EXTERNA EM PANDEMIA

Q5. Nesse mesmo ano letivo, a avaliação externa foi cancelada em praticamente todos os anos. Na sua opinião, considera que o cancelamento da <u>avaliação externa para fins de avaliação sumativa</u> <u>e classificação</u> foi uma boa decisão?

DESBLOQUEADORES (se necessário):

- Q5.1 Na sua opinião, quem foi mais beneficiado com essa alteração?
- Q5.2 Considera que essa decisão permitiu diminuir as desigualdades entre os alunos?
- Q5.3 Na sua opinião, professores, alunos e pais ficaram satisfeitos com estas alterações?

Q6. Considera importante a existência de avaliação externa das aprendizagens?

DESBLOQUEADORES (se necessário):

Q6.1 Na sua opinião, para que servem as avaliações externas?

Q6.2 Na sua opinião, qual é a principal utilidade das avaliações externas?

Q6.3 Na sua opinião, as avaliações externas permitem conhecer o que os alunos sabem e são capazes de fazer e o que é que se faz com essa informação?

Q6.4 Na sua opinião, as avaliações externas são (ou não) importantes para o processo de ensino/aprendizagem?

Q7. Considera que as avaliações internas e as avaliações externas são consistentes?

DESBLOQUEADORES (se necessário):

Q7.1 A nível nacional?

Q7.2 E na sua escola?

Q8. Na sua opinião, as avaliações externas influenciam as práticas de avaliação sumativa interna e classificação feita pelos professores?

Q9. Nestes últimos anos não houve avaliação externa para fins de classificação, apenas para o acesso ao ensino superior. Na sua opinião, essa alteração influenciou as práticas de avaliação sumativa e de classificação dos professores?

BLOCO 3: FUTURO DA AVALIAÇÃO:

Q10. Considera que o cancelamento da avaliação externa para fins de avaliação sumativa e classificação deveria continuar no futuro? Porquê?

Q11. Considera que deve ser estabelecido um novo equilíbrio entre avaliação interna e externa das aprendizagens? Que equilíbrio?

- DESBLOQUEADORES (se necessário):
- Q11.1 Que provas deveriam existir?
- Q11.2 Em que anos deveriam existir provas?
- Q11.3 Que tipo de provas e com que peso?
- Q11.4 Para que fins?

Q12. Recentemente o Ministério da Educação anunciou que as provas de aferição e os exames nacionais vão passar a ser exclusivamente digitais. Essa informação está já a ter algum impacto na organização da escola?

DESBLOQUEADORES (se necessário):

- Q11.1 Já têm todas as necessidades ao nível recursos materiais?
- Q11.2 Facilita a gestão de recursos humanos durante os exames?
- Q11.3 Vão fazer testes (internos) em computador?
- Q11.4 O assunto já foi debatido em conselho pedagógico?
- Q11.5 Concorda com a realização da avaliação externa em formato digital?

Q13. Por coincidência, durante a pandemia foi implementado o projeto MAIA, um projeto de formação de professores para melhorar as práticas de avaliação. A sua escola participou nesse projeto?

Q14. Considera que a pandemia (**se sim em Q13:** e o projeto MAIA) contribuíram para a mudança de culturas e práticas de avaliar as aprendizagens nesta escola?

SE SIM em Q14:

Q15. Que diferenças existem, na atualidade, em termos de avaliação das aprendizagens, relativamente ao pré-pandemia?

DESBLOQUEADORES (se necessário):

Q15.1 Em relação ao pré-pandemia, considera que existe uma preocupação diferente relativamente à avaliação das aprendizagens?

Q15.2 Em relação ao pré-pandemia, considera que existe novas práticas e instrumentos de avaliação nesta escola?

Q15.3 Em relação ao que existia pré-pandemia, existem novos critérios de avaliação?

Q15.4 Em relação ao que existia pré-pandemia, existem novas ponderações nos critérios de avaliação?

Q15.5 Em relação ao que existia pré-pandemia, considera que é usada uma maior diversidade de instrumentos de avaliação?

Q15.6 Essas diferenças devem-se sobretudo à pandemia ou à formação através do projeto MAIA?

Q16. Para o futuro, que alterações consideraria pertinente introduzir ao nível da avaliação sumativa e classificação?

Q17. Existe alguma coisa, relativamente ao impacto da pandemia nos processos de avaliação das aprendizagens, internos ou externos, que considere pertinente referir e que não foi falada?

CONCLUSÃO DA ENTREVISTA

- 1. Agradecer, uma vez mais, a entrevista.
- 2. Debriefing.
- 3. Sair.

APPENDIX B Teachers' questionnaire

O presente estudo surge no âmbito de um projeto de investigação a decorrer no Iscte – Instituto Universitário de Lisboa, financiado pela Fundação para a Ciência e Tecnologia (FCT) com a referência 2020.05847.BD. O estudo tem por objetivo conhecer a perceção dos professores quanto às implicações do ensino a distância nas suas práticas de avaliação das aprendizagens. Pretende-se, ainda, conhecer a perceção dos professores relativamente às políticas de avaliação das aprendizagens promovidas pelo Ministério de Educação durante a pandemia Covid-19. O estudo é realizado pelo investigador Gabriel Cipriano (gabriel.cipriano@iscte-iul.pt), que poderá contactar caso pretenda esclarecer uma dúvida ou partilhar algum comentário. A sua participação no estudo consiste em preencher um inquérito por questionário online, com duração estimada de 10 minutos. Não existem riscos associados à sua participação no estudo, sendo a sua participação estritamente **voluntária**. Para além de voluntária, a sua participação é também **anónima** e **confidencial**. Os dados obtidos serão sujeitos apenas a tratamento estatístico e nenhuma resposta será analisada ou reportada individualmente. A divulgação de dados será feita apenas para efeitos de ensino, comunicação em encontros e/ou publicações científicas.

Declaro ter compreendido os objetivos de quanto me foi proposto, ter-me sido dada oportunidade de fazer todas as perguntas sobre o presente estudo e para todas elas ter obtido resposta esclarecedora, pelo que...

- SIM, aceito participar no estudo. (1)
- NÃO, não quero participar no estudo. (2)

Q1. Qual o seu grupo de recrutamento?

▼ 110 - 1º Ciclo (1) ... Outro não especificado (36)

Q2. Qual a sua habilitação académica mais elevada?

▼ Bacharelato ou equivalente (1) ... Doutoramento (5)

Q3. A que nível ou níveis de ensino leciona neste ano letivo?

- 1º Ciclo (1)
- 2º Ciclo (2)
- 3º Ciclo (3)
- Secundário Cursos Científico-Humanísticos (4)
- Secundário Cursos Profissionais (5)
- Outro (6)

Q4. No presente ano letivo, leciona no ensino público ou no ensino privado?

▼ Ensino público (1) ... No ensino público e privado (3)

Q5. A escola pública onde leciona tem uma das seguintes características administrativas?

▼ TEIP - Território Educativo de Intervenção Prioritária (1) ... Não sei (5)

	Nunca (1)	Poucas vezes (2)	Muitas vezes (3)	Sempre (4)
Ferramentas para a criação de recursos				
educativos digitais (p.ex.,Padlet©, Kahoot©,				
Quizziz©, PiktoChart©) (1)	0	\bigcirc	\bigcirc	0
Vídeos (2)	0	\bigcirc	\bigcirc	\bigcirc
Plataformas de ensino (3)	0	\bigcirc	\bigcirc	\bigcirc
Páginas da internet (4)	0	0	\bigcirc	\bigcirc
Manuais da disciplina (5)	0	\bigcirc	\bigcirc	\bigcirc
Apontamentos de sua autoria (6)	0	\bigcirc	\bigcirc	\bigcirc
PowerPoint© (7)	0	\bigcirc	\bigcirc	\bigcirc
Fichas de trabalho (8)	0	\bigcirc	\bigcirc	\bigcirc

Q6. Na atualidade, com que frequência utiliza os seguintes suportes em sala de aula?

	Nunca (1)	Poucas vezes (2)	Muitas vezes (3)	Sempre (4)
Trabalho de projeto (1)	0	0	0	0
Partilha e discussão dos produtos finais entre pares (2)	0	\bigcirc	0	0
Jogos e desafios (3)	0	\bigcirc	0	\bigcirc
Trabalho por portefólio (4)	0	\bigcirc	\bigcirc	\bigcirc
Testes escritos em papel (5)	0	\bigcirc	\bigcirc	\bigcirc
Testes digitais (6)	0	\bigcirc	\bigcirc	\bigcirc
Apresentações orais (8)	0	\bigcirc	\bigcirc	\bigcirc
Apoio ao desenvolvimento de técnicas e competências de estudo (9)	0	\bigcirc	\bigcirc	\bigcirc
Planificação e calendarização das tarefas com os alunos (10)	0	\bigcirc	\bigcirc	\bigcirc

Q7. Na atualidade, com que frequência utiliza as seguintes metodologias ou abordagens?

Q8. Comparando com as suas práticas antes da pandemia, classifique o seu grau de concordâncias com as seguintes afirmações:

	Discordo totalmente (1)	Discordo parcialmente (2)	Nem concordo nem discordo (3)	Concordo parcialme nte (4)	Concordo totalmente (5)
Hoje em dia diversifico mais os meus instrumentos de avaliação. (1)	0	0	0	0	0
Hoje em dia promovo mais o trabalho autónomo dos alunos e a auto-regulação das aprendizagens. (2)	0	\bigcirc	0	0	0
Hoje em dia faço aulas com suporte online para alunos que estão doentes e não podem vir à escola. (3)	0	\bigcirc	0	0	\bigcirc
Hoje em dia comunico e forneço feedback aos meus alunos através de e-mail ou plataformas digitais. (4)	0	\bigcirc	\bigcirc	0	\bigcirc
Hoje em dia uso chamadas por video- conferência para comunicar ou reunir com colegas, alunos e/ou pais. (5)	0	\bigcirc	\bigcirc	0	\bigcirc
Hoje em dia preocupo-me mais com a avaliação formativa e em melhorar as aprendizagens dos meus alunos. (6)	0	\bigcirc	\bigcirc	0	\bigcirc
Hoje em dia preocupo-me mais com a avaliação sumativa e a classificação dos meus alunos. (7)	0	\bigcirc	0	\bigcirc	0

Q9. Para si, o sucesso educativo de um aluno depende sobretudo de:

(Arraste as frases para ordenar por importância)

_____ Do Ministério da Educação e das políticas de educação (1)

_____ Da escola e da sua organização interna (2)

_____ Dos professores (3)

- _____ Dos próprios alunos (4)
- _____ Dos encarregados de educação (5)

Q10. Para si, a retenção de um aluno depende sobretudo de:

(Arraste as frases para ordenar por importância)

- _____ Do Ministério da Educação e das políticas de educação (1)
- _____ Da escola e da sua organização interna (2)
- _____ Dos professores (3)
- _____ Dos próprios alunos (4)
- _____ Dos encarregados de educação (5)

Q11. Relativamente à retenção escolar, indique o seu grau de concordância com as seguintes afirmações:

	Discordo totalmente (1)	Discordo parcialmente (2)	Nem concordo nem discordo (3)	Concordo parcialmente (4)	Concordo totalmente (5)
Os processos de avaliação definidos pelo Ministério da Educação são adequados ao que se faz nas escolas. (1)	0	0	0	0	0
A retenção de um aluno deve ocorrer apenas no final de um ciclo de ensino. (2)	0	\bigcirc	\bigcirc	\bigcirc	0
A retenção de um aluno é feita considerando o que é melhor para ele. (3)	0	\bigcirc	\bigcirc	\bigcirc	\bigcirc
Faço uso da retenção como medida educativa. (4)	0	\bigcirc	\bigcirc	\bigcirc	\bigcirc
A retenção é benéfica para os alunos. (5)	0	\bigcirc	\bigcirc	\bigcirc	0
Para determinados alunos a retenção é inevitável. (6)	0	\bigcirc	\bigcirc	\bigcirc	\bigcirc

Q12. Durante os anos da pandemia Covid-19, o Ministério da Educação decidiu cancelar todas as provas de aferição e todas as provas finais de ciclo (9º ano). Os exames nacionais do ensino secundário foram realizados apenas como provas de acesso ao ensino superior. Considera ter sido uma boa decisão? *(selecione a(s) frase(s) com as quais concorda)*

- O cancelamento das provas de aferição do <u>2º ano</u> foi uma boa decisão. (1)
- O cancelamento das provas de aferição do <u>5º ano</u> foi uma boa decisão. (2)
- O cancelamento das provas de aferição do <u>8º ano</u> foi uma boa decisão. (3)
- O cancelamento das provas finais de ciclo do <u>9º ano</u> foi uma boa decisão. (4)
- O cancelamento dos exames nacionais do <u>11º e 12º anos</u> para conclusão do ensino secundário foi uma boa decisão. (5)

Q13. No pós-pandemia, indique em que ciclo(s) de ensino considera importante existir provas de aferição:

- No 1º Ciclo. (1)
- No 2º Ciclo. (2)
- No 3º Ciclo. (3)
- No Secundário. (4)
- 😣 Não deviam existir provas de aferição. (5)

Q14. No pós-pandemia, indique em que ciclo(s) de ensino considera importante existir exames nacionais:

- No 1º Ciclo. (1)
- No 2º Ciclo. (2)
- No 3º Ciclo. (3)
- No Secundário. (4)
- 🛞 Não deviam existir exames nacionais. (5)

Q15. Relativamente à conclusão do ensino secundário, selecione aquela que é, para si, a melhor opção:

- Os exames nacionais devem existir apenas para o acesso ao ensino superior. (1)
- Os exames nacionais devem existir <u>tanto para a co</u>nclusão do ensino secundário como para o acesso ao ensino superior. (2)
- Os exames nacionais devem existir <u>apenas para a conclusão do ensino secundário</u> e as instituições de ensino superior deviam desenvolver o seu próprio modelo de seleção e acesso. (3)
- Os exames nacionais <u>deviam ser completamente abolidos</u> e as instituições de ensino superior deviam desenvolver o seu próprio modelo de seleção e acesso. (4)

Q16. Relativamente à desmaterialização da avaliação externa, selecione a(s) frase(s) com as quais concorda.

- Concordo com a desmaterialização das provas de aferição do 2º ano. (1)
- Concordo com a desmaterialização das provas de aferição do 5º ano. (2)
- Concordo com a desmaterialização das provas de aferição do 8º ano. (3)
- Concordo com a desmaterialização das provas finais de ciclo do 9º ano. (4)
- Concordo com a desmaterialização dos exames nacionais do ensino secundário. (5)
- 🛞 Não concordo com a desmaterialização da avaliação externa. (6)

Q17. Desde Setembro de 2019, pouco antes do início da pandemia Covid-19, que está a ser desenvolvido e implementado o Projeto MAIA - Monitorização, Acompanhamento e Investigação em Avaliação Pedagógica. Participou em alguma formação no âmbito do projeto MAIA?

- Sim (1)
- Não (2)

Apresentar esta pergunta: If Participou em alguma formação no âmbito do projeto MAIA? = Sim

Q18. Relativamente ao projeto MAIA, selecione as afirmações com as quais concorda:

- O projeto MAIA foi importante para refletir sobre as minhas práticas de avaliação. (1)
- O projeto MAIA foi importante para melhorar as minhas práticas de avaliação. (2)
- O projeto MAIA trouxe inovações às minhas práticas de avaliação. (3)
- Na minha atividade docente aplico novos procedimentos que aprendi no âmbito do projeto MAIA. (4)
- Com o projeto MAIA passei a preocupar-me mais com a avaliação formativa. (5)
- O projeto MAIA aumentou o volume de trabalho que tenho em tarefas relacionadas com a avaliação. (6)
- \otimes O projeto MAIA não teve qualquer impacto nas minhas práticas de avaliação. (7)

Apresentar esta pergunta: If Participou em alguma formação no âmbito do projeto MAIA? = Não

Q19. Relativamente ao projeto MAIA, selecione as afirmações com as quais concorda:

- Ainda não participei em nenhuma formação do projeto MAIA mas tenho interesse em participar. (1)
- Tenho interesse em participar nas formações do projeto MAIA mas tenho receio que aumente o meu volume de trabalho. (2)
- Ainda não participei em nenhuma formação do projeto MAIA porque não sinto necessidade em fazer formação ao nível da avaliação. (3)
- Tenho uma imagem negativa sobre o projeto MAIA e não tenciono participar nas formações. (4)
- 🛞 Não concordo com nenhuma das anteriores afirmações. (5)

Para terminar, por favor, indique:

Q20. Quantos anos completos de serviço tem?

▼ 0 (1) ... 50 (51)

Q21. Qual a sua idade?

▼ 18 (1) ... 70 (53)

Q22. Qual o seu sexo?

- Feminino (1)
- Masculino (2)
- Outro / prefiro não dizer (3)

Q23. Em que região do país leciona neste ano letivo?

▼ Alentejo (10) ... Não sei (21)

Apresentar esta pergunta: If Em que região do país leciona neste ano letivo? = Norte

Q24a. Norte Mais concretamente, em que zona da região Norte leciona?

▼ Alto Minho (1) ... Não sei (9)

Apresentar esta pergunta: If Em que região do país leciona neste ano letivo? = Centro

Q24b. Centro Mais concretamente, em que zona da região Centro leciona?

▼ Beira Baixa (1) ... Não sei (2)

Apresentar esta pergunta:If Em que região do país leciona neste ano letivo? = Alentejo

Q24c. Alentejo Mais concretamente, em que zona do Alentejo leciona?

▼ Alentejo Central (1) ... Não sei (3)

Q25. Existe alguma coisa relativamente ao impacto da pandemia nos processos de avaliação das aprendizagens, internos ou externos, que considere pertinente referir e que não foi abordado?

- Não. (1)
- Sim. Por favor, indique o que pretende referir: (2)

Agradecemos o tempo dedicado a responder a este inquérito por questionário.

A sua resposta foi gravada.

APPENDIX C

IAVE's president interview script

ANTES DE COMEÇAR

- 1. Agradecer a disponibilidade e voltar a enquadrar o estudo.
- 2. Garantir que consentimento informado foi assinado, solicitar autorização para a gravação do áudio da entrevista.
- 3. Permitir a desistência a qualquer momento.

PARTE A: FORMATO DAS PROVAS

Q1. Estreitamento do currículo:

Nas provas em papel, práticas laboratoriais não são testadas, oralidade não é testada; há toda uma série de coisas do currículo que não são possíveis de testar. Pelo seu formato digital, não se corre o risco de prejudicar a escrita, privilegiando determinados domínios do currículo passíveis de serem testados digitalmente, estreitando, ainda mais o currículo? Ou acredita que isso é mitigado pelo incluir de novas opções como simulações, som e vídeo?

Q2. Limitações nas práticas de avaliação interna:

Atualmente, em termos de avaliação interna, existe uma cultura de teste escrito, baseada nos exames nacionais. Testes semelhantes, nos itens, na classificação, na duração, etc.

Inexistência de práticas internas de ensino e avaliação digital (apenas formativa).

Profs têm incapacidade de produzir "testes" semelhantes aos "e-exames".

Vários autores salientam que é importante que o candidato esteja familiarizado com o modelo de testagem e que tenham oportunidade de aprendizagem. **Como capacitar os professores e alunos para isso? Nomeadamente a escrita matemática, a Física, e Geometria Descritiva? Os centros de formação e as editoras estão a ser envolvidas neste processo?**

Q3. Acessibilidade e universalidade da prova:

Nas escolas, é praticamente inexistente práticas de avaliação interna digital. Para além disso, os alunos do 2º ano não dominam a motricidade fina, e TIC não faz parte da matriz curricular do 2º ano. Muitos alunos não têm se quer computador pessoal e uma boa parte dos alunos não tem literacia digital.

3.1 Não corremos o risco de causar injustiça na avaliação, não por não dominarem o construto testado, mas por não dominarem o digital?

3.2 Não se corre o risco de favorecer contextos privilegiados (ex. privados) e prejudicar os desfavorecidos (ex. TEIP), acentuando desigualdades?

Q4. Contextos de aprendizagem:

Uma crítica dos diretores à avaliação externa é a falta de integração dos contextos de aprendizagem no <u>desenho</u> e <u>interpretação</u> dos resultados, em particular nas PA. Esforços têm sido feitos para integrar indicadores socioeconómicos na leitura de resultados, nomeadamente no PISA e TIMS; mas raramente os contextos de aprendizagem são considerados em Portugal, no desenho e leitura dos resultados das provas. Fazem-se acomodações, mas o construto testado é sempre o mesmo. Outros diretores referem ainda que a existência de avaliação externa limita a flexibilização curricular e a adoção de medidas diferenciadas nas escolas, pois há um "programa a cumprir".

Está previsto no futuro, para as PA, PFC e EN, integrar os contextos de aprendizagem na <u>conceção e</u> <u>leitura</u> dos resultados?

Q5. Modelos alternativos:

O teste (a testagem) externo em larga escala, baseado em psicométricos, tem vantagens ao nível da validade e fiabilidade, princípios essenciais para a análise estatística e são uma forma "fácil" para ter informação do sistema. Porém, não é isento de críticas. Vários autores referem que a existência de avaliação externa na forma de teste, especialmente os exames pelos seus impactos, promove uma cultura interna de teste com consequências negativas. As PFC e os EN vão ser reintroduzidos apenas com pequenas alterações nas ponderações.

Modelos alternativos de avaliação externa podem passar por retirar as consequências dos testes externos (como as nossas PA), a realização de testes adaptativos em função das respostas dos alunos em perguntas anteriores, ou fazer avaliação externa sem recorrer à testagem, com entrevistas, portefólios ou observações, por exemplo.

Estamos a caminhar para um destes paradigmas num futuro próximo?

<u>PARTE B – IMPLEMENTAÇÃO</u>

Q6. Pré-testes às provas digitais:

Desde 2018 foram feitos testes piloto para a implementação do DAVE. O IAVE diz que têm resultados semelhantes em papel ou online, mas a divulgação extensiva desses resultados é quase inexistente, não havendo referência às dificuldades na implementação. Tudo isto aparece apenas numa nota de 1 página no relatório das PA de 2022. Nas minhas entrevistas com diretores escolares que foram feitas entre outubro e fevereiro, alguns relataram uma total falta de informação por parte de ME, JNE e IAVE. Após as minhas entrevistas, sei que, entretanto, o IAVE e o JNE fizeram reuniões em Jan, Fev, e Mar de 2023 com os diretores de escola, bem como Webinares. Em todo o caso, o Guia para a realização de provas digitais apenas saiu em abril de 2023. Por que motivo não existiu mais informação mais cedo?

Q7. Limitações Infraestruturas:

A maioria dos diretores que entrevistei fala em equipamento informático insuficiente e obsoleto, para além da falta de internet nas escolas. Isso foi muito evidente nas entrevistas zoom que realizei, cheias de falhas. Existe, de facto, uma desigualdade entre escolas em termos de equipamento. Na página web do IAVE, é referido que *"em aplicação piloto, as provas de aferição em suporte digital (PAD) foram realizadas numa amostra de escolas definida pelo IAVE. Esta amostra de escolas, sendo de conveniência, foi constituída por escolas que consideraram ter capacidade tecnológica para os alunos poderem realizar as provas em suporte eletrónico." (IAVE, 2022).* Após as reuniões com os diretores realizadas em Jan, Fev e Mar; foram feitos investimentos às condições do parque escolar em articulação com a Direção Geral dos Estabelecimentos Escolares? As limitações das infraestruturas tecnológicas e digitais já foram ultrapassadas?

Q8. Segurança das provas:

Com um parque informático obsoleto, alguns diretores realçaram dúvidas quanto à segurança das provas, com a inexistência de um exame em papel como garantia de que a prova foi realizada pelo(s) aluno(s). Com o armazenamento da informação recolhida digitalmente a ser guardada em servidores escolares sujeitos a ataques informático, nomeadamente na solução *standalone*, é possível garantir que uma prova digital é tão segura como uma prova em papel?

Q9. Limitações RH:

Muitos diretores referiram também a falta de técnicos de informática nas escolas, com uma classe docente com limitações nas TIC; que faz um fraco uso das TIC nas práticas letivas; a pouca *e-avaliação* existente nas escolas é sobretudo feita com caracter formativo/lúdico e não sumativo. Está previsto um investimento adicional ao nível dos RH em articulação com DGRHE, DGE, JNE? (formação / novas contratações)

Q10. Operacionalização e implementação:

A associação de professores de TIC foi muito crítica às primeiras provas de aferição digitais. Membros do JNE colocaram lugares à disposição. As restantes provas de aferição, aparentemente, pelo quse ouviu na comunicação social, correram bem. Haverá, certamente, muitas lições a tirar desta 1ª fase. Q10.1 Que alterações estão já previstas serem feitas em função das PA deste ano letivo? Q10.2 As PAD deste ano permitem avançar com segurança para as PFC no próximo ano e depois os EN, já com consequências nos percursos escolares dos alunos?

Q11. Condições de aplicação PFC e EN:

As PAD foram feitas em 2 turnos para aumentar os recursos disponíveis. Para as PFC e os EN, será adotado o mesmo sistema em turnos? Pelo facto de terem consequências, requer algum cuidado especial em relação às PAD?

Q12. Finalização da entrevista:

Alguma coisa que não foi abordada e considera pertinente referir?

CONCLUSÃO DA ENTREVISTA

- 1. Agradecer, uma vez mais, a entrevista.
- 2. Debriefing.
- 3. Sair.

APPENDIX D



CONSENTIMENTO INFORMADO

O presente estudo surge no âmbito de um projeto de investigação a decorrer no **Iscte** – **Instituto Universitário de Lisboa**, plano de trabalhos esse financiado pela Fundação para a Ciência e Tecnologia (FCT) com a referência 2020.05847.BD.

O estudo tem por objetivo conhecer a forma como as escolas lidaram com a avaliação das aprendizagens durante a pandemia Covid-19. Pretende-se, ainda, conhecer as implicações do ensino remoto de emergência nas políticas e práticas de avaliação das aprendizagens nas escolas portuguesas.

A sua participação no estudo, que será muito valorizada pois irá contribuir para o avanço do conhecimento neste domínio da ciência, consiste em responder a 16 questões numa entrevista através de videoconferência, que deverá durar cerca de 30 minutos.

O Iscte – Instituto Universitário de Lisboa é o responsável pelo tratamento dos seus dados pessoais, recolhidos e tratados exclusivamente para as finalidades do estudo, tendo como base legal o seu consentimento informado, de acordo com o previsto no art. 6º, nº1, alínea a) do Regulamento Geral de Proteção de Dados.

O estudo é realizado por Gabriel Cipriano (*gabriel.cipriano@iscte-iul.pt*), que poderá contactar caso pretenda esclarecer uma dúvida, partilhar algum comentário ou exercer os seus direitos relativos ao tratamento dos seus dados pessoais. Poderá utilizar o contacto indicado para solicitar o acesso, a retificação, o apagamento ou a limitação do tratamento dos seus dados pessoais.

A participação neste estudo é **confidencial**. Os seus dados pessoais serão tratados apenas pelo investigador Gabriel Cipriano, vinculado ao dever de sigilo e confidencialidade. O Iscte garante a utilização das técnicas, medidas organizativas e de segurança adequadas para proteger as informações pessoais. É exigido a todos os investigadores que mantenham os dados pessoais confidenciais.

ISCTE-IUL Instituto Universitário de Lisboa 🖂 Av. das Forças Armadas, 1649-026 Lisboa 📞 351 217 903 000 www.iscte-iul.ot_www.facebook.com//SCTEIUL twitter.com/iscteiul_www.linkedin.com/comoanv/iscteiul_www.flickr.com/abotos/iscteiul_www.voutube.com/user/jultv





Além de confidencial, a participação no estudo é estritamente **voluntária**: pode escolher livremente participar ou não participar. Se tiver escolhido participar, pode interromper a participação e retirar o consentimento para o tratamento dos seus dados pessoais em qualquer momento, sem ter de prestar qualquer justificação. A retirada de consentimento não afeta a legalidade dos tratamentos anteriormente efetuados com base no consentimento prestado.

Os seus dados pessoais serão conservados por um período inferior a 1 mês, o tempo previsto para a transcrição e anonimização da entrevista. Após este período, o registo áudio e vídeo serão eliminados, garantindo-se o seu anonimato nos resultados do estudo, apenas divulgados para efeitos estatísticos, de ensino, comunicação em encontros ou publicações científicas.

Não existem riscos expectáveis associados à participação no estudo.

O lscte não divulga ou partilha com terceiros a informação relativa aos seus dados pessoais. O lscte tem um Encarregado de Proteção de Dados, contactável através do email <u>dpo@iscte-iul.pt.</u> Caso considere necessário, tem ainda o direito de apresentar reclamação à autoridade de controlo competente – Comissão Nacional de Proteção de Dados.

Declaro ter compreendido os objetivos de quanto me foi proposto e explicado pelo investigador Gabriel Cipriano, ter-me sido dada oportunidade de fazer todas as perguntas sobre o presente estudo e para todas elas ter obtido resposta esclarecedora. **Aceito** participar no estudo e consinto que os meus dados pessoais sejam utilizados de acordo com a informações que me foram disponibilizadas.

Sim 🛛 Não 🗖

Assinatura:						
www.iscte-iul.pt	ISCTE-IUL I www.facebook.com/ISC	nstituto Universitário de Lisb TEIUL twitter.com/iscteiul w	oa 🖂 Av. das Forças Arma ww.linkedin.com/company	idas, 1649-026 Lisboa 📞 351 217 /iscte-iul www.flickr.com/photos/	903 000 'iscteiul www.youtube.i	com/user/iultv
A3ES Apércia de Asalação e Acreditação do Emiro Superior	AACSB		Several and Severa	CONTRACT AND	Project Management Institute	Townshind Constant

NOTA: Assinatura dos relatores intencionalmente omitida para proteção dos seus dados pessoais.



This page intentionally left blank

APPENDIX E

Mann-Whitney U tests from Chapter 6

Sample A vs Sample B	Ν	Mann- Whitney U test	Z	<i>p</i> value
1.1. Grade repetition should only occur by the end of a cycle.				
Whole samples test	4 506	2 532 929.5	1.923	.055
Subgroup tests				
Female subgroup	3 381	1 422 429.5	2.208	.027
Male subgroup	1 073	147 618.0	0.819	.413
≤ 39 years old subgroup	323	13 497.0	0.601	.548
Between 40-49 years old subgroup	1 577	307 642.5	0.654	.513
≥ 50 years old subgroup	2 606	832 355.0	1.269	.204
ISCED 5 subgroup	76	751.5	0.602	.547
ISCED 6 subgroup	3 439	1 478 667.0	1.769	.077
ISCED 7 or 8 subgroup	991	120 145.5	0.731	.465
Public school subgroup	4 270	2 292 183.0	2.030	.042
Private school subgroup	236	5 816.0	0.471	.638
1 st cycle (grades 1-4) subgroup	898	112 349.5	3.671	< .001
2 st cycle (grades 5-6) subgroup	724	58 859.0	-1.670	.095
3 rd cycle / Secondary (grades 7-12) subgroup	2640	851 063.0	0.692	.489

Sample A vs Sample B.

Sample A vs Sample B	Ν	Mann- Whitney U test	Z	<i>p</i> value
1.2. For some students grade repetition is inevitable.				
Whole samples test	4 505	2 433 975.5	-0.479	.632
Subgroup tests				
Female subgroup	3 382	1 344 206.5	-0.791	.429
Male subgroup	1 071	140 875.5	-0.466	.641
≤ 39 years old subgroup	323	12 135.5	-1.151	.250
Between 40-49 years old subgroup	1 578	308 031.0	0.655	.512
≥ 50 years old subgroup	2 604	801 711.5	-0.374	.708
ISCED 5 subgroup	76	640.5	-0.675	.499
ISCED 6 subgroup	3 437	1 407 583.0	-0.793	.428
ISCED 7 or 8 subgroup	992	121 159.5	0.922	.357
Public school subgroup	4 269	2 198 180.0	-0.434	.664
Private school subgroup	236	5 476.5	-0.300	.764
1 st cycle (grades 1-4) subgroup	899	97 303.5	-0.580	.562
2 st cycle (grades 5-6) subgroup	722	62 846.5	-0.077	.939
3 rd cycle / Secondary (grades 7-12) subgroup	2639	835 307.5	-0.143	.887

		Mann-		
Sample A vs Sample B	N	Whitney U	Z	<i>p</i> value
		test		
1.3. Grade repetition is beneficial for students.				
Whole samples test	4 507	2 521 378.5	1.593	.111
Subgroup tests				
Female subgroup	3 382	1 402 879.5	1.418	.156
Male subgroup	1 072	145 351.0	0.400	.689
≤ 39 years old subgroup	324	12 993.0	-0.154	.877
Between 40-49 years old subgroup	1 577	317 880.5	1.831	.067
≥ 50 years old subgroup	2 606	833 912.5	1.347	.178
ISCED 5 subgroup	76	648.0	-0.536	.592
ISCED 6 subgroup	3 437	1 471 079.0	1.523	.128
ISCED 7 or 8 subgroup	994	122 038.5	1.021	.307
Public school subgroup	4 271	2 274 037.0	1.505	.132
Private school subgroup	236	5 836.0	0.500	.617
1 st cycle (grades 1-4) subgroup	899	99 206.5	-0.060	.952
2 st cycle (grades 5-6) subgroup	723	65 484.0	0.858	.391
3 rd cycle / Secondary (grades 7-12) subgroup	2640	881 374.5	2.317	.020

Sample A vs Sample B	Ν	Mann- Whitney U test	Z	<i>p</i> value
1.4. I use grade repetition as an educational measure				
Whole samples test	4 499	2 352 454.5	-2.290	.022
Subgroup tests				
Female subgroup	3 375	1 341 000.0	-0.718	.473
Male subgroup	1 071	128 533.0	-2.955	.003
≤ 39 years old subgroup	323	12 056.0	-1.198	.231
Between 40-49 years old subgroup	1 575	285 515.0	-1.848	.065
≥ 50 years old subgroup	2 601	788 992.0	-0.987	.324
ISCED 5 subgroup	76	763.5	0.719	.472
ISCED 6 subgroup	3 430	1 365 765.5	-2.101	.036
ISCED 7 or 8 subgroup	993	113 165.5	-1.007	.314
Public school subgroup	4 265	2 124 563.5	-2.240	.025
Private school subgroup	234	5 209.0	-0.737	.461
1 st cycle (grades 1-4) subgroup	900	92 294.0	-1.946	.052
2 st cycle (grades 5-6) subgroup	723	61 107.0	-0.778	.437
3 rd cycle / Secondary (grades 7-12) subgroup	2633	814 358.0	-1.094	.274

		Mann-		
Sample A vs Sample B_0	Ν	Whitney U	Z	<i>p</i> value
		test		
1.1. Grade repetition should only occur by the end of a cycle.				
Whole samples test	3 539	1 569 301.0	0.226	.821
Subgroup tests				
Female subgroup	2 629	869 182.0	0.292	.770
Male subgroup	875	95 333.5	0.703	.482
≤ 39 years old subgroup	298	11 387.0	0.611	.541
Between 40-49 years old subgroup	1 297	211 977.0	0.278	.781
≥ 50 years old subgroup	1 944	471 246.0	0.009	.993
ISCED 5 subgroup	66	511.5	0.565	.572
ISCED 6 subgroup	2 702	925 815.5	0.778	.437
ISCED 7 or 8 subgroup	771	71 426.5	-0.987	.324
Public school subgroup	3 351	1 409 676.5	0.425	.671
Private school subgroup	188	4 052.0	0.080	.937
1 st cycle (grades 1-4) subgroup	710	67 638.0	2.153	.031
2 st cycle (grades 5-6) subgroup	576	38 063.5	-1.743	.081
3 rd cycle / Secondary (grades 7-12) subgroup	2 056	521 820.5	-0.465	.642

Sample A vs Sample B₀.

		Mann-		
Sample A vs Sample B ₀	N	Whitney U	Z	<i>p</i> value
		test		
1.2. For some students grade repetition is inevitable.				
Whole samples test	3 540	1 580 845.5	0.597	.550
Subgroup tests				
Female subgroup	2 631	874 772.0	0.525	.599
Male subgroup	874	911 120.5	-0.420	.674
≤ 39 years old subgroup	299	10 135.0	-1.326	.185
Between 40-49 years old subgroup	1 297	213 754.0	0.556	.578
≥ 50 years old subgroup	1 944	481 122.0	0.850	.395
ISCED 5 subgroup	66	445.5	-0.423	.672
ISCED 6 subgroup	2 701	909 360.5	-0.042	.967
ISCED 7 or 8 subgroup	773	79 829.5	1.740	.082
Public school subgroup	3 352	1 411 895.5	0.477	.633
Private school subgroup	188	4 132.0	0.318	.750
1 st cycle (grades 1-4) subgroup	711	63 783.5	0.504	.614
2 st cycle (grades 5-6) subgroup	575	40 909.0	-0.203	.839
3 rd cycle / Secondary (grades 7-12) subgroup	2 056	536 849.0	0.718	.473

		Mann-		
Sample A vs Sample B ₀	N	Whitney U	Z	<i>p</i> value
		test		<u>.</u>
1.3. Grade repetition is beneficial for students.				
Whole samples test	3 541	1 643 814.5	2.698	.007
Subgroup tests				
Female subgroup	2 631	920 253.0	2.935	.003
Male subgroup	874	92 975.0	0.106	.915
≤ 39 years old subgroup	299	10 985.0	-0.098	.922
Between 40-49 years old subgroup	1 297	223 105.0	1.983	.047
≥ 50 years old subgroup	1 945	495 863.0	2.023	.043
ISCED 5 subgroup	66	436.5	-0.509	.610
ISCED 6 subgroup	2 702	954 789.0	2.248	.025
ISCED 7 or 8 subgroup	773	80 604.0	1.970	.049
Public school subgroup	3 353	1 466 235.0	2.444	.015
Private school subgroup	188	4 406.0	1.113	.266
1 st cycle (grades 1-4) subgroup	711	65 555.0	1.168	.243
2 st cycle (grades 5-6) subgroup	575	43 220.0	1.003	.316
3 rd cycle / Secondary (grades 7-12) subgroup	2 057	569 451.0	3.173	.002

Sample A vs Sample B ₀	N	Mann- Whitney U	Z	p value
		test		•
1.4. I use grade repetition as an educational measure				
Whole samples test	3 533	1 516 833.0	-1.371	.170
Subgroup tests				
Female subgroup	2 624	857 189.0	-0.175	.861
Male subgroup	873	85 358.5	-1.949	.051
≤ 39 years old subgroup	298	9 773.5	-1.666	.096
Between 40-49 years old subgroup	1 294	199 783.0	-1.443	.149
≥ 50 years old subgroup	1 941	467 473.5	-0.179	.858
ISCED 5 subgroup	66	547.5	1.067	.286
ISCED 6 subgroup	2 695	880 256.0	-1.306	.191
ISCED 7 or 8 subgroup	772	72 618.5	-0.623	.533
Public school subgroup	3 346	1 355 132.0	-1.430	.153
Private school subgroup	187	3 894.5	-0.285	.776
1 st cycle (grades 1-4) subgroup	711	59 898.5	-0.975	.330
2 st cycle (grades 5-6) subgroup	574	41 023.0	-0.063	.950
3 rd cycle / Secondary (grades 7-12) subgroup	2 051	514 353.5	-0.823	.410
		Mann-		
--	-------	-----------	--------	----------------
Sample A vs Sample B ₁	Ν	Whitney U	Z	<i>p</i> value
		test		
1.1. Grade repetition should only occur by the end of a cycle.				
Whole samples test	2 811	963 628.5	3.701	< .001
Subgroup tests				
Female subgroup	2 083	553 247.5	4.208	< .001
Male subgroup	711	52 284.5	.639	.523
≤ 39 years old subgroup	190	2 110.0	0.205	.838
Between 40-49 years old subgroup	937	95 665.5	1.047	.295
≥ 50 years old subgroup	1 684	361 109.0	2.428	.015
ISCED 5 subgroup	55	240.0	0.349	.727
ISCED 6 subgroup	2 147	552 851.5	2.563	.010
ISCED 7 or 8 subgroup	609	48 719.0	2.977	.003
Public school subgroup	2 697	882 506.5	3.587	< .001
Private school subgroup	114	1 764.0	1.108	.268
1 st cycle (grades 1-4) subgroup	581	44 711.5	4.449	< .001
2 st cycle (grades 5-6) subgroup	444	20 795.5	-0.894	.371
3 rd cycle / Secondary (grades 7-12) subgroup	1 648	329 242.5	2.109	.035

Sample A vs Sample B₁.

		Mann-		
Sample A vs Sample B_1	N	Whitney U	Z	p value
		test		
1.2. For some students grade repetition is inevitable.				
Whole samples test	2 809	853 130.0	-1.892	.058
Subgroup tests				
Female subgroup	2 082	469 434.5	-2.437	.015
Male subgroup	710	49 755.0	-0.332	.740
≤ 39 years old subgroup	189	2 000.5	0.089	.929
Between 40-49 years old subgroup	938	94 277.0	0.554	.580
≥ 50 years old subgroup	1 682	320 589.5	-1.794	.073
ISCED 5 subgroup	55	195.0	-0.743	.458
ISCED 6 subgroup	2 146	498 222.5	-1.604	.109
ISCED 7 or 8 subgroup	608	41 330.0	-0.635	.525
Public school subgroup	2 695	786 284.5	-1.596	.111
Private school subgroup	114	1 344.5	-1.444	.149
1 st cycle (grades 1-4) subgroup	581	33 520.0	-1.901	.057
2 st cycle (grades 5-6) subgroup	443	21 937.5	0.151	.880
3 rd cycle / Secondary (grades 7-12) subgroup	1 647	298 458.5	-1.338	.181

		Mann-		
Sample A vs Sample B ₁	Ν	Whitney U	Z	<i>p</i> value
		test		
1.3. Grade repetition is beneficial for students.				
Whole samples test	2 810	877 564.0	-0.663	.507
Subgroup tests				
Female subgroup	2 092	482 626.5	-1.352	.176
Male subgroup	711	52 376.0	0.666	.506
≤ 39 years old subgroup	190	2 008.0	-0.223	.824
Between 40-49 years old subgroup	937	94 775.5	0.769	.442
≥ 50 years old subgroup	1 683	338 049.5	0.029	.976
ISCED 5 subgroup	55	211.5	-0.302	.763
ISCED 6 subgroup	2 145	516 290.0	-0.144	.886
ISCED 7 or 8 subgroup	610	41 434.5	-0.765	.445
Public school subgroup	2 696	807 802.0	-0.448	.654
Private school subgroup	114	1 430.0	-0.915	.360
1 st cycle (grades 1-4) subgroup	581	33 651.5	-1.799	.072
2 st cycle (grades 5-6) subgroup	444	22 264.0	0.294	.769
3 rd cycle / Secondary (grades 7-12) subgroup	1 647	311 923.5	0.198	.843

Sample A vs Sample B ₁	N	Mann- Whitney U test	Z	p value
1.4. I use grade repetition as an educational measure				
Whole samples test	2 810	835 621.5	-2.775	.006
Subgroup tests				
Female subgroup	2 082	483 811.0	-1.253	.210
Male subgroup	711	43 174.5	-3.179	.001
≤ 39 years old subgroup	190	2 282.5	0.881	.378
Between 40-49 years old subgroup	938	85 732.0	-1.776	.076
≥ 50 years old subgroup	1 682	321 518.5	-1.676	.094
ISCED 5 subgroup	55	216.0	-0.205	.838
ISCED 6 subgroup	2 145	485 509.5	-2.469	.014
ISCED 7 or 8 subgroup	610	40 547.0	-1.204	.229
Public school subgroup	2 697	769 431.5	-2.556	.011
Private school subgroup	113	1 314.5	-1.419	.156
1 st cycle (grades 1-4) subgroup	582	32 395.5	-2.567	.010
2 st cycle (grades 5-6) subgroup	445	20 084.0	-1.583	.113
3 rd cycle / Secondary (grades 7-12) subgroup	1 646	300 004.5	-1.076	.282

Sample B_0 vs Sample B_1	Ν	Mann- Whitney U test	Z	p value
1.1. Grade repetition should only occur by the end of a cycle.				
Whole samples test	2 662	886 486.0	3.677	< .001
Subgroup tests				
Female subgroup	2 050	539 110.5	4.146	< .001
Male subgroup	560	36 134.0	0.168	.866
≤ 39 years old subgroup	158	1 653.5	-0.046	.963
Between 40-49 years old subgroup	920	92 591.0	0.863	.388
≥ 50 years old subgroup	1 584	327 071.5	2.528	.011
ISCED 5 subgroup	31	106.0	0.044	.965
ISCED 6 subgroup	2 029	500 061.0	1.977	.048
ISCED 7 or 8 subgroup	602	50 108.5	4.126	< .001
Public school subgroup	2 492	778 355.0	3.354	< .001
Private school subgroup	170	3 328.0	1.479	.139
1 st cycle (grades 1-4) subgroup	505	33 823.0	2.690	.007
2 st cycle (grades 5-6) subgroup	428	21 453.0	0.623	.533
3 rd cycle / Secondary (grades 7-12) subgroup	1 576	311 582.0	2.637	.008

Sample B₀ vs Sample B₁.

Sample B_0 vs Sample B_1	Ν	Mann- Whitney U	Z	<i>p</i> value
		test	_	p
1.2. For some students grade repetition is inevitable.				
Whole samples test	2 661	771 981.5	-2.563	.010
Subgroup tests				
Female subgroup	2 051	449 671.5	-3.145	.002
Male subgroup	558	35 596.5	0.022	.983
≤ 39 years old subgroup	158	1 769.5	0.832	.405
Between 40-49 years old subgroup	921	90 287.5	0.105	.916
≥ 50 years old subgroup	1 582	280 774.5	-2.745	.006
ISCED 5 subgroup	31	98.0	-0.322	.747
ISCED 6 subgroup	2 027	454 767.0	-1.692	.091
ISCED 7 or 8 subgroup	603	37 723.0	-2.208	.027
Public school subgroup	2 491	686 850.0	-2.120	.034
Private school subgroup	170	2 368.5	-2.046	.041
1 st cycle (grades 1-4) subgroup	506	26 207.5	-2.441	0.15
2 st cycle (grades 5-6) subgroup	426	20 948.5	0.385	.700
3 rd cycle / Secondary (grades 7-12) subgroup	1 575	271 744.0	-2.110	.035

		Mann-		
Sample B_0 vs Sample B_1	Ν	Whitney U	Z	<i>p</i> value
		test		
1.3. Grade repetition is beneficial for students.				
Whole samples test	2 663	764 848.0	-2.990	.003
Subgroup tests				
Female subgroup	2 051	439 168.0	-3.953	< .001
Male subgroup	559	36 889.0	0.652	.514
≤ 39 years old subgroup	159	1634.5	-0.203	.840
Between 40-49 years old subgroup	920	85 767.0	-1.090	.276
≥ 50 years old subgroup	1 584	290 083.0	-1.725	.084
ISCED 5 subgroup	31	109.0	0.174	.861
ISCED 6 subgroup	2 027	449 758.0	-2.060	.039
ISCED 7 or 8 subgroup	605	37 344.5	-2.552	.011
Public school subgroup	2493	680 463.0	-2.548	.011
Private school subgroup	170	2 336.5	-2.142	.032
1 st cycle (grades 1-4) subgroup	506	25 364.0	-3.018	.003
2 st cycle (grades 5-6) subgroup	427	20 051.5	-0.508	.612
3 rd cycle / Secondary (grades 7-12) subgroup	1 576	268 547.0	-2.496	.013

Sample B_0 vs Sample B_1	Ν	Mann- Whitney U test	Z	<i>p</i> value
1.4. I use grade repetition as an educational measure				
Whole samples test	2 655	783 574.5	-1.743	.081
Subgroup tests				
Female subgroup	2 044	471 161.5	-1.149	.250
Male subgroup	558	32 366.0	-1.842	.065
≤ 39 years old subgroup	158	2 021.0	1.758	.079
Between 40-49 years old subgroup	918	87 191.5	-0.639	.523
≥ 50 years old subgroup	1 579	289 567.5	-1.582	.114
ISCED 5 subgroup	31	85.0	-0.864	.388
ISCED 6 subgroup	2 020	454 346.0	-1.458	.145
ISCED 7 or 8 subgroup	604	40 784.5	-0.766	.444
Public school subgroup	2 487	696 108.0	-1.451	.147
Private school subgroup	168	2 486.5	-1.292	.196
1 st cycle (grades 1-4) subgroup	507	27 305.5	-1.767	.077
2 st cycle (grades 5-6) subgroup	427	18 852.0	-1.571	.116
3 rd cycle / Secondary (grades 7-12) subgroup	1 569	283 921.0	-0.391	.696