



Examining the effects of the COVID-19 pandemic and professional development on teachers' grade repetition beliefs and practice

Gabriel Cipriano¹ · Susana da Cruz Martins²

Received: 8 June 2024 / Accepted: 18 August 2025
© The Author(s) 2025

Abstract

Grade repetition has been a controversial educational decision consisting of requiring students to remain at the same grade level for an additional school year. During the COVID-19 pandemic, in some countries where grade repetition rates are usually high, a notable decrease in these rates was observed. This study investigates the impact of the COVID-19 pandemic on teachers' grade repetition beliefs and practices. Additionally, it examines the effect of professional development in assessment literacy on these beliefs and practices. Using data from two independent samples collected through two survey questionnaires before and after the COVID-19 pandemic, where a portion of surveyed teachers participated in an assessment literacy professional development programme, we found that while the pandemic had minimal influence on teachers' grade repetition beliefs, professional development had a significant effect: Grade repetition beliefs were not affected by the COVID-19 pandemic, although the trend towards reducing the practice of grade repetition was positively influenced by the COVID-19 pandemic. Hence, both grade repetition beliefs and practices were positively influenced by professional development in assessment literacy. The results are consistent with the literature and institutional statistical indicators.

Keywords Grade repetition beliefs · Assessment literacy · COVID-19 pandemic · Teachers' professional development

1 Introduction

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,

✉ Gabriel Cipriano
gabriel.cipriano@iscte-iul.pt
Susana da Cruz Martins
susana.martins@iscte-iul.pt

¹ Centro de Investigação e Estudos de Sociologia (CIES-Iscte), Iscte – IUL, Lisbon, Portugal

² Centro de Investigação e Estudos de Sociologia (CIES-Iscte), Iscte – IUL, Lisbon, Portugal

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 Emergency Remote Teaching (ERT), unequal 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). For these reasons, during this period, many countries decided to introduce exceptional regulations to internal and external assessments, grading, and grade transition conditions (e.g. CNE, 2021; 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 dropped considerably 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 countries analysed 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).

The pandemic introduced unprecedented challenges to teachers’ assessment practices, prompting exceptional measures that questioned traditional criteria for grade promotion and repetition. Despite the long-standing debate on the effectiveness of grade repetition, only limited research has explored how disruptions such as the COVID-19 pandemic might influence teachers’ beliefs about grade repetition. Thus, the aim of the present research is to understand if the COVID-19 pandemic affected teachers’ grade repetition beliefs, and whether the decline in grade repetition rates, particularly notable in the 2019/2020 school year, are complemented by a change in teachers’ grade repetition beliefs. It also aims to study the interaction effect of professional development on assessment literacy with teachers’ grade repetition beliefs and practice. To this end, Portugal was taken as a case study and two independent samples collected through two survey questionnaires were compared. The first sample was collected before the COVID-19 pandemic. The second sample was collected after the COVID-19 pandemic, when a number of these teachers participated in an assessment literacy’ professional development programme during the pandemic years. It is hoped that this article will contribute usefully to the field of educational assessment and teachers’ professional development, by presenting the evolution of some beliefs about grade repetition and its practice, as well as the influence of the COVID-19 pandemic and professional development on these beliefs.

1.1 Grade repetition

In response to 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* form of 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 a form of educational intervention is highly controversial. When students held back do not receive extra educational services during the repeated year, studies have shown that grade repetition itself becomes an ineffective tool to deal with underachievers (Nunes et al., 2018), for it has been shown that repeating a school year alone has negative effects on academic achievement, with socio-emotional and behavioural consequences for students, increasing the likelihood of premature dropping out from school (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’ progress (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 native language, i.e. students with an immigrant background and/or a different mother tongue (European Commission et al., 2020; OECD, 2023b). When dealing with cultural and linguistic diversity, Wiese and 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. In spite of this, Costa et al. (2015) report that in Portugal the introduction of changes to previous teaching practices has encountered some resistance. Teachers’ concerns about complying the syllabus and with exams are often obstacles to the development of different teaching strategies and activities to improve learning.

1.2 The Portuguese education system and grade repetition policy

The Portuguese compulsory education system is divided into basic and secondary education. Basic education is composed of three cycles: the first cycle includes grades 1 to 4 (lower ISCED 1); the second cycle includes grades 5 and 6 (upper ISCED 1); the third cycle includes grades 7 to 9 (ISCED 2); secondary education includes grades 10 to 12 (ISCED 3).

Since at least 1992, Portuguese educational legislation has stated that formative assessment should prevail in classrooms at all grade levels. Teachers are the main agents in the assessment process, and they are also fully responsible for evaluating and certifying students in grade levels without external assessments (Fernandes, 2009). In these circumstances, the teacher’s role is not limited to instruction; their identity now also includes the role of “assessor” (Xu & Brown, 2016).

Despite the emphasis in policy that the teacher’s role as an assessor should prioritize formative assessment, there remains a gap between what has been legally stipulated and current assessment practices in most Portuguese schools. Both Portuguese and international research suggest that classroom assessments are predominantly used to collect information for grading purposes, rather than to support student improvement or help them overcome learning difficulties (Fernandes, 2009). The 2012 OECD *Reviews of Evaluation*

and Assessment in Education: Portugal (Santiago et al., 2012) confirmed what many studies had already suggested: while formative assessment is strong in policy, it remains under-implemented in Portuguese classroom practice, with significant consequences for pedagogy and learning. According to the OECD team, in practice, the Portuguese system often prioritises summative approaches, placing excessive emphasis on results and data. Although this reflects a results-driven culture, it undermines the formative and developmental purposes of assessment. Furthermore, the persistently high levels of grade repetition in Portugal—among the highest in the OECD—point to a system that does not consistently place the learner at the centre. In its report, the OECD recommended a gradual reduction of grade repetition, to be accompanied by alternative support measures for struggling students and ongoing professional development for teachers. For such reforms to be effective, teachers would require sustained support to strengthen their assessment literacy and develop inclusive, student-centred classroom strategies that reduce dependence on grade repetition (Santiago et al., 2012).

In 2017, in line with this OECD recommendation, the *Students' Profile by the End of Compulsory Schooling (PASEO)* was approved (see Martins et al., 2017). The PASEO serves as a reference document for the organisation of the entire education system, contributing to the coherence and alignment of decisions across various dimensions of curriculum development. The purpose of this document is to support the organisation and management of curricula, as well as the definition of strategies, methodologies, and pedagogical-didactic procedures to be used in the practice of teaching.

In order to implement the PASEO, Decree-Law No. 55/2018 was approved in 2018. This Decree-Law outlines the curriculum, guiding principles, and assessment processes for basic and secondary education, reaffirming that formative assessment should prevail in classrooms at all grade levels. It also emphasizes that assessment should be an integral part of the teaching and learning process, with the main objective of improving both teachers' approaches and students' learning. This Portuguese legislation further stipulates that grade repetition in basic education should preferably occur at the end of a cycle of studies (i.e. in grades 4, 6, or 9), and only under exceptional circumstances during the middle of a cycle. As a consequence, grade repetition is not allowed at the end of grade 1. In addition, for students facing grade repetition, teachers are required to define appropriate support measures to ensure access to the curriculum. This includes the adoption of alternative teaching and learning strategies, as well as the mobilisation of suitable educational resources to support the full development of student learning (Decree-Law No. 55/2018).

Despite these recent developments in policy, as shown by Cipriano and Martins (2021), Portuguese teachers have their own conceptions and beliefs about assessment purposes and practice, and these beliefs often superimpose themselves on recommendations and legislation regarding assessment and grading. Accordingly, analysing DGEEC (2023) institutional data, it is possible to observe that grade repetition remains a common practice at all grade levels, and grade repetition rates have been consistently higher at the beginning of a cycle of studies in basic education (i.e. on grades 2, 5 and 7), contrary to the Decree-Law No. 55/2018 legal framework that proposes the end of a cycle (i.e. on grades 4, 6 and 9). Cavaco et al (2021) also highlight that, when compared with other European countries, grade repetition rates in Portugal are higher among male students. However, Portuguese teachers tend to reject the association between the use of grade repetition and students' sex.

1.3 Assessment and grading during the COVID-19 pandemic in Portugal

The closure of Portuguese schools during the COVID-19 pandemic affected students' learning unevenly, the worst affected being pupils from the most vulnerable families in socioeconomic and cultural terms (Torres & Alves, 2024). Despite relevant institutional and pedagogical responses to deal with the COVID-19 pandemic in Portugal (e.g. Flores e Gago, 2020), teachers underline the lack of appropriate and clear guidance from the Government throughout the crisis, especially regarding critical components of the pedagogical process, such as assessment, and feedback (Seabra et al., 2021). Additionally, the study conducted by the *Conselho Nacional de Educação* [National Education Council] (CNE, 2021) also revealed that the ERT raised teachers' doubts and questions about assessment methods, instruments, criteria, objects, and trust; in which approximately 82% of teachers that participated in their survey indicated that assessment of learning was "difficult" or even "very difficult" to perform. According to the CNE, Portuguese teachers' practices regarding assessment, if not their conceptions, were challenged by the change in circumstances. Given that assessment remains an area in need of conceptual and practical consolidation within teachers' professional development, the CNE highlights the diversity of solutions adopted and the experiences developed—both because they were based on the search for tools to gather evidence of student learning, and because they fostered a context favourable to reflection on the purposes and methods of assessment (CNE, 2021). Flores et al. (2021) also underscore a tension around assessment and feedback on students' remote homework, with doubts regarding the authorship of these works and the validity and reliability of these assessments. Consequently, 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), they 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 dropped considerably in Portugal during the 2019/2020, to the lowest grade repetition rate ever recorded in Portugal (e.g. Cipriano & Martins, 2023). Such a shift in the use of grade repetition by teachers as sole assessors, in the context of both institutional pressures to ensure student assessments and teachers' struggles to evaluate learning fairly under emergency conditions, leads us to the following Research Question (RQ):

RQ1: As the COVID-19 pandemic affected teachers' grade repetition practices, how did the pandemic impact teachers' grade repetition beliefs?

1.4 Assessment literacy and professional development during the pandemic

For many years, teachers' assessment literacy was primarily associated with standardised measurement and classroom testing (Pastore & Andrade, 2019), as it is essential for teachers to understand how these tests work and their implications for instructional quality (Popham, 2011). However, assessment literacy is increasingly being recognized as an integral part of teacher professionalism (Xu & Brown, 2016), and over time, evidence has shown a gradual shift in conceptions of assessment literacy (DeLuca et al., 2015). Triggered by the extensive review by Black and Wiliam (1998), which emerged as a seminal work that played a vital role in redefining the concept, nowadays, assessment literacy refers to the knowledge, skills, and dispositions required to design, administer, interpret, and use

assessments effectively, within the classroom context and the school system, to support student learning and make informed instructional decisions anchored in educational standards (DeLuca et al., 2015; Pastore & Andrade, 2019). In other words, it encompasses an individual's understanding of the fundamental assessment concepts and procedures that are likely to influence educational decisions (Popham, 2011). Thus, assessment literacy is seen as an important factor in how teachers use information about student learning to teach more effectively and respond to students' learning needs (Pastore & Andrade, 2019).

According to Tomchin and 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. An approach of this kind, recognizing the importance of research, theory, and practice, can encourage authentic change with positive outcomes for students. Training programmes aimed at changing teachers' beliefs may 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, with a view to improving Portuguese teachers' assessment literacy and students' learning through teachers' professional development, the national project for Training, Supervision and Research in Classroom Assessment (the MAIA Project) was created. The MAIA project was a nationwide project with voluntary participation that began in September 2019, a few months before the COVID-19 lockdowns. The project was designed and developed in terms of six angles: 1) the theoretical and foundations angle, 2) the conceptual angle, 3) the training angle, 4) the follow-up angle, 5) the monitoring angle, and 6) the research angle. When schools closed due to the COVID-19 pandemic in March 2020, it was decided to continue the MAIA project with distance training (Fernandes, 2021a, 2021b). This led to the decision to test whether the participation in this professional development programme affected Portuguese teachers grade repetition beliefs, prompting the following research question:

RQ2: How did engagement with the MAIA project affect teachers' understanding of their role as assessor and thus their grade repetition beliefs?

1.5 Concept clarification for the present 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 et al., 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.

According to Crahay et al. (2010), teachers' beliefs serve four main functions: 1) making sense of experiences lived; 2) providing an identity function through the sharing of common beliefs; 3) guiding and regulating practices and behaviours; and 4) serving a justificatory or self-defensive role. 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, grade repetition 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. Consequently, teachers' grade repetition beliefs can serve as an interpretive and guiding framework through which they mediate the uptake and implementation of theoretical knowledge (Xu & Brown, 2016).

From a sociological perspective, grade repetition beliefs 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 made more complex 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). Yet the focus of empirical research in this area appears to have been on examining the gap between teachers' frequency of use of an assessment practice and their beliefs about the importance of that practice (Barnes et al., 2015). In this sense, this article aims to provide interpretive perspectives on the influence of public programmes or policies, such as the MAIA Project, on grade repetition beliefs and practices, with the possibility of adaptation and recreation by school policies (Fernandes, 2021a, 2021b).

2 Methods

Beliefs about grade repetition are not directly observable, and they can only be inferred from responses or indicators (Eagly & Chaiken, 1993). To understand how the COVID-19 pandemic affected Portuguese teachers' grade repetition beliefs and practice, a quantitative approach was implemented 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 in April 2019, before the COVID-19 pandemic and before the beginning of the MAIA project, *versus* Sample B, collected in a post-pandemic period, in October 2023.

2.1 Databases

2.1.1 Sample A

The database created by Cipriano and 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 carried out in April 2019, aimed to understand the conceptions and beliefs of Portuguese teachers about educational 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.

To form sample A, from the Cipriano and 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*; and Item 1.4. *I use grade repetition as an educational measure*. On the four items, a ten-point Likert scale was used, with verbal descriptors at the extremes, where 1 means "Strongly disagree" and 10 means "Strongly agree".

To understand the latent structure underlying this set of four items (Tabachnick & Fidell, 2014), and to establish the validity of these four items to accurately measure the intended construct (Groves et al., 2004), a factor analysis with varimax rotation was conducted ($KMO = .68$, Bartlett test $\chi^2(6) = 1219.964$, $p < .001$), 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 and their practice. 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 & Watson, 1995). Note that for internal consistency analysis, item 1.1. was inverted, as this item had a negative loading in the factor analysis.

2.1.2 Sample B

To form sample B, a survey questionnaire was prepared to be conducted with the same procedures as those used to collect sample A, replicating the previously selected four items. The selection of items used previously in other studies made it possible to compare data from two time periods with two independent samples (Sudman & Bradburn, 1982). Although Sample A was collected using a 10-point scale with descriptors at the endpoints, all items in Sample B used a five-point Likert scale. This allowed responses to be given within the intended dimensions using verbal labels (Sudman & Bradburn, 1982), where 1 means "Strongly disagree" and 5 means "Strongly agree". Using verbal descriptors for each category clarifies the inherent social meaning of each option (Calapez et al., 2016), while still allowing for comparison after the recoding of Sample A. Furthermore, teachers in the survey were also asked if they had participated in the MAIA training project to understand the effect of the interaction of these four items with participation in the MAIA project. Lastly, 10 additional items were also added in terms of age, sex, region, and qualification, as well as other professional backgrounds to characterise the sample.

2.1.3 Sampling processes for data collection

Data supporting the Cipriano and Martins (2021) study, with $n = 1844$ 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. 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 was also 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 to head teachers were sent during the last week of September 2023, asking them to forward the survey through all the other teachers in their schools and to voluntarily participate in the study in the course of the month of October 2023.

As these two samples under study were formed through nonprobabilistic processes, by performing the Chi-squared goodness-of-fit test, it was possible to observe sampling bias regarding sex, age, qualification, school sector and teaching level. However, it should be noted that there is no pattern or predictability in these two samples (Stopher, 2012), 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 1.

2.2 Participants

Both in the samples and in the population, Table 1 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 approximately 60% of teachers work with students in lower and/or upper secondary education (grades 7–12).

Table 1 Teachers from the compulsory school system by sex, age, qualification, school sector and teaching level

	Sample A (<i>n</i> = 1 844)		Sample B (<i>n</i> = 2 673)		Population ^{a)} (<i>N</i> = 124 066)	
	<i>n</i>	Valid %	<i>n</i>	Valid %	<i>N</i>	%
Sex						
Female	1 331	72.2	2 058	78.6	93 999	75.8
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)}						
1st cycle (grades 1–4)	393	22.4	508	20.2	28 608	23.1
2nd cycle (grades 5–6)	296	16.9	429	17.0	21 613	17.4
3rd cycle / Secondary (grades 7–12)	1 064	60.7	1 581	62.8	50 221	59.5

Notes: Sample A was collected in 2019, in the pre COVID-19 pandemic. Sample B was collected in 2023, in the post-COVID-19 pandemic. 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

2.3 Measures

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 $\chi^2(6) = 1420.581$, $p < .001$). 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 & 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 in the factor analysis.

In Sample B, 1 703 teachers (63.7%) reported that they did not participate in the MAIA project (defining subsample B₀), and 970 teachers (36.3%) reported having participated in the MAIA training project (defining subsample B₁).

2.4 Procedures and data analysis

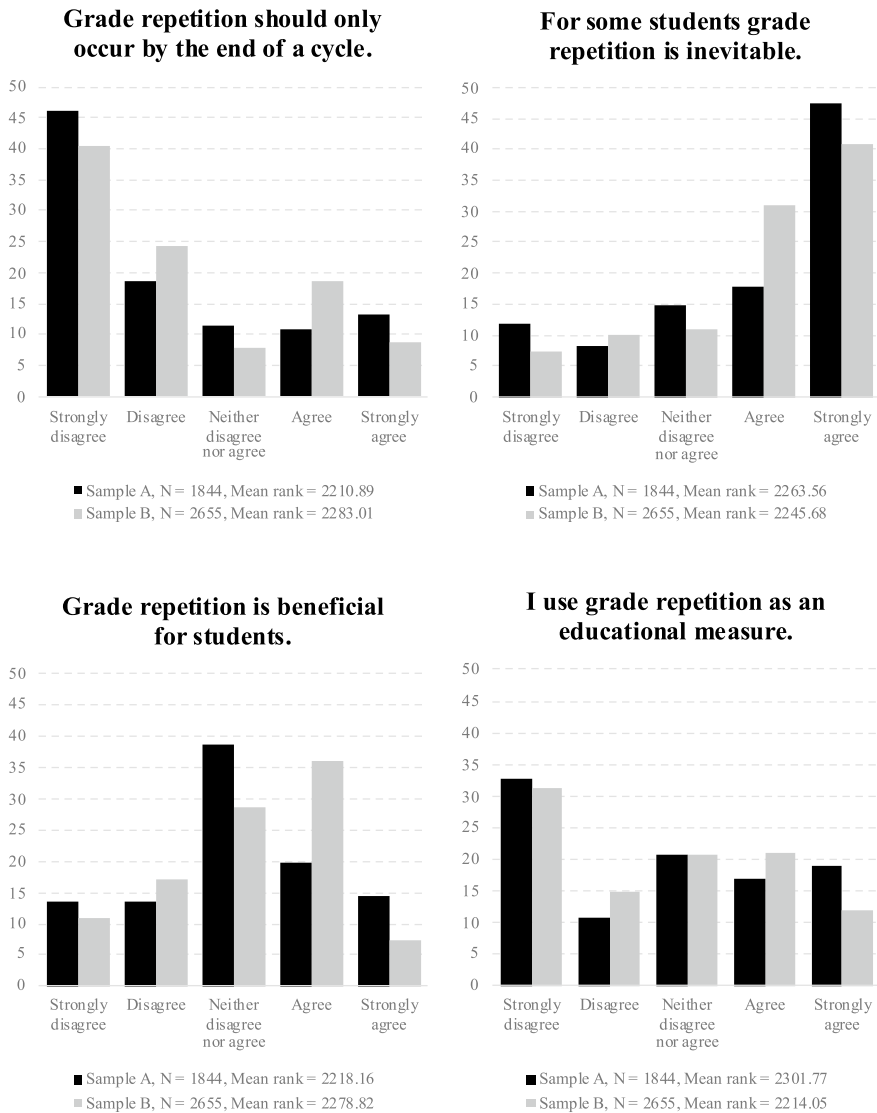
To allow comparisons between the two independent samples A and B, the samples were compiled into a single SPSS (2022 version) file. Before its compilation, items answered with a ten-point Likert scale in Sample A were recoded into a five-point Likert scale, as in Sample B. Considering that sampling bias 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 were performed with these variables as covariates, using complementary log–log, negative log–log, and probit functions according to class distributions (Marôco, 2021). However, as in all models the data did not fit the model well and the slope homogeneity was rejected for all the items (item 1.1. $\chi^2_{PL}(27) = 131.807$, $p < .001$; item 1.2. $\chi^2_{PL}(27) = 157.858$, $p < .001$; item 1.3. $\chi^2_{PL}(27) = 249.490$, $p < .001$; item 1.4. $\chi^2_{PL}(27) = 105.211$, $p < .001$), logistic regressions were only used as exploratory analysis.

Alternatively, Mann–Whitney U tests were performed to compare the distributions of the two samples A and B. To understand the effect of participation on the MAIA project, the very same procedures were 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 were also performed selecting the categories of these five variables as subgroups to be tested.

3 Results

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 Fig. 1 and Table 2.



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.

Fig. 1 Teachers' grade repetition beliefs and practice by sample (%). 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

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

Table 2 Distributions comparisons between samples and subsamples

	N	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. Sources: Cipriano & Martins, 2021; Primary data, 2023

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 using 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 2 and Table 3, 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. Thus, 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.

3.2 Understanding the effect of participation on the MAIA project

In the previous section, Table 2 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 make it possible to understand the effects of participation in the MAIA professional development project on teachers' grade repetition beliefs, as interpreted in the following points.

3.2.1 Grade repetition should only occur at the end of a cycle

When testing sample A vs sample B and sample A vs subsample B_0 , 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 B_1 and subsample B_0 vs subsample B_1 , it was possible to observe that significant changes have occurred, as shown in Table 2. 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.

3.2.2 For some students, grade repetition is inevitable.

When testing sample A vs sample B, sample A vs subsample B_0 , and sample A vs subsample B_1 , it is possible to observe that no significant changes occurred pre- and post-COVID-19, as also shown in Table 2. 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 occurred. This means that teachers that did not participate on the MAIA project comparing to the teachers that did participate believe more that for some students, grade repetition is inevitable.

3.2.3 Grade repetition is beneficial for students.

When testing sample A vs sample B, and sample A vs subsample B_1 , it is possible to observe that no changes occurred before and after the COVID-19 pandemic for teachers that participated in the MAIA project, as shown in Table 2. However, when testing sample A vs subsample B_0 and subsample B_0 vs subsample B_1 , this belief about grade repetition benefits increased for teachers that did not participate in the MAIA project. Thus, the COVID-19 pandemic in interaction with the nonparticipation in the MAIA project, increased the belief that grade repetition is beneficial for students.

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 took place, 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 2. 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 in the MAIA project, but it decreased more for teachers that did take part.

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. In Table 3, the subgroups where the Mann–Whitney U tests had significant changes are reported.

Table 3 Distributions comparisons within subgroups

	N	Mann–Whitney U	Z value	p value	Effect size r
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
1st 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					
3rd 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-COVID-19, in October 2023. Subsample B₀ represents teachers that did not participated in the MAIA project. Subsample B₁ represents teachers that participated in the MAIA project. Sources: Cipriano & Martins, 2021; Primary data, 2023

Within these subgroup tests with significant changes, it seems to be relevant to highlight two of them when considering teaching level: the first cycle teachers' subgroup, where regardless of their participation in the MAIA project, increased the belief that *grade repetition should only occur by the end of a cycle*, as shown in Table 3; 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 did not change in the post-COVID-19 pandemic period. However, it should be noted once again that all these effect sizes are very small (Cohen, 1992).

4 Discussion

Regarding the use of grade repetition as an educational intervention, when comparing the pre-COVID-19 period (Sample A) with the post-COVID-19 period (Sample B), teachers in 2023 reported a decrease in the use of grade repetition, continuing the previous trend of its reduction (DGEEC, 2023; OECD, 2023b). Although the effect size is very small, the differences between 2019 and 2023 are statistically significant. We believe that the reported reduction in the use of grade repetition can be attributed to the scientific evidence regarding its inefficiency (Allen et al., 2009; Jimerson, 2001a, 2001b; Nunes et al., 2018; Santos et al., 2023), the international pressure for changes in grade repetition practices (European Commission et al., 2020; OECD, 2023a, 2023b; Santiago et al., 2012), as well as national pressure resulting from recent changes in assessment policies (e.g. Martins et al., 2017; Decree-Law 54/2018; Decree-Law 55/2018). This finding aligns with Portuguese institutional statistical indicators, which reflect this trend over the last few decades (DGEEC, 2023). Moreover, it seems that the COVID-19 pandemic has had a positive impact on this trend (OECD, 2023b). In Portugal, grade repetition rates saw a considerable decrease across all grade levels during the 2019/2020 school year (Cipriano & Martins, 2023), reaching their lowest value ever recorded. According to DGEEC data (2023), this trend continued in subsequent years, aligning with the pre-existing downward trajectory.

Regarding RQ1, when examining 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 at the end of a cycle. This fact is in line with literature reporting that teachers' beliefs tend to self-perpetuate over time (Pajares, 1992), and they are deeply rooted in school culture (Goldring, 2002; Santana, 2019). Note that the contextual challenges posed by the COVID-19 pandemic, such as the lack of clear guidance on assessment practices (Seabra et al., 2021), the difficulties inherent in remote assessment (CNE, 2021), the tensions reported by teachers regarding assessment validity, reliability, and fairness (Flores et al., 2021; Sandvik et al., 2023), and the additional responsibility placed on teachers in their decision-making (Cooper et al., 2022), undeniably pressured teachers to adapt their grade repetition practices. However, the extent to which these circumstances also triggered changes in underlying beliefs remains questionable. Yet based on this study, it seems plausible that the

pandemic context generated situational changes in practice without fundamentally transforming belief systems. Furthermore, these findings suggest that adaptations were reactive and context-specific.

In fact, according to Crahay et al. (2010), few researchers believe that it is possible to change teachers' beliefs in the short term. Nevertheless, despite the great difficulty in making teachers' beliefs evolve, teachers do change. When teachers participate in voluntary reflective and collaborative professional development programmes with adequate support and follow-up, they can experience dramatic changes, such as shifts in orientations and beliefs (e.g. Le Fevre & Richardson, 2002; Richardson & Placier, 2001). Programmes designed to improve assessment literacy encourage teachers to examine strategies for developing and implementing effective classroom practices (Tomchin & Impara, 1992). Moreover, these programmes help ensure that an assessment literate teacher understands and differentiates the aims of assessment and is aware of both the negative effects of grade retention and alternative strategies. In this way, an assessment literate teacher can make better use of assessment information to draw inferences about student learning, redefine classroom practices, and make more informed decisions (Pastore & Andrade, 2019), including those related to grade repetition. Regarding RQ2, these considerations are also supported in our study when analysing data from sample B by subgroups (B_0 and B_1), i.e. taking into account 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. Those who did participate showed positive changes in their beliefs about grade repetition, expressing less belief that it is beneficial for students and less belief that it is inevitable for some students. Furthermore, they demonstrated greater support for the idea that grade repetition should occur only at the end of an educational cycle. 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 is worth noting that a teacher may need to compromise their assessment beliefs when faced with different circumstances (Xu & Brown, 2016). This may partly explain the change in grade repetition practices and beliefs among teachers participating in the MAIA project. Even so, this study clearly suggests that while emergency circumstances may accelerate temporary adjustments in practices, lasting changes in beliefs likely require sustained reflection, support, and professional development opportunities beyond crisis contexts.

From the results section, it should also be highlighted that in the post-COVID-19 context, 1st cycle teachers increased their belief that grade repetition should happen at 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 heterogeneity 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 (DGECC, 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 held back in a previous grade level. This also suggests that the level of education at which one teaches

influences the evolution of one's beliefs about grade retention. However, further research on this topic should be undertaken to confirm these hypotheses.

5 Conclusions

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 at an early stage in a teacher's 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 in spite of disruptive events, such as the COVID-19 pandemic lockdowns and the ERT.

While it is true that there was a drop in grade repetition rates in Portugal during the 2019/2020 school year, this seems to be related to doubts and questions about assessment methods, instruments, criteria, objects, and trust (CNE, 2021), rather than to a shift in teachers' grade repetition beliefs. Without appropriate and clear guidance for 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.

Nevertheless, even with very small effect sizes, this 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 in changing teachers' beliefs and practices than disruptive events such as the COVID-19 pandemic and the ERT.

Acknowledgements This work was supported by Fundação para a Ciência e Tecnologia (FCT) [the Portuguese National Funding Agency for Science, Research and Technology] under the grants 2020.05847.BD and R&D Unit UIDB/03126/2020.

Funding Open access funding provided by FCTIFCCN (b-on). Fundação para a Ciência e a Tecnologia, 2020.05847.BD, Gabriel Cipriano, UIDB/03126/2020

Data Availability Data will be made available on reasonable request.

Declarations

Conflict of interest The authors report there are no competing interests to declare, and all authors have approved the final article.

Ethical approval In accordance with ethical standards, appropriate informed consent was provided to all research participants.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- 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. <https://doi.org/10.3102/0162373709352239>
- 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*, Routledge. (pp. 284–300).
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Calapez, T., Ramos, M., & Botelho, M. C. (2016). Describe all or just a few? The influence of verbal descriptors on likert-type variables on items' and scales' distribution. *Sociologia Online*, 12, 87–109.
- 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.
- 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. *Education Research and Policy Practice*, 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.
- Cipriano, G., & Martins, S. C. (2023). Avaliação das aprendizagens em Portugal durante a pandemia [Assessment of learning in Portugal during the pandemic]. In D. Fonseca, A. Neto-Mendes, M. Gonçalves, A. Ventura, & J. A. Costa (Eds.), *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.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3), 309–319. <https://doi.org/10.1037/1040-3590.7.3.309>
- CNE. (2021). *Educação em tempo de pandemia: Problemas, respostas e desafios das escolas* [Education in pandemic times: problems, responses, and challenges of schools]. Conselho Nacional de Educação.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- 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*, 29(5), 534–554. <https://doi.org/10.1080/0969594X.2022.2069084>
- Costa, A. F., Pegado, E., Ávila, P., & Coelho, A. R. (2015). Evaluating the Portuguese National reading plan: teachers' perceptions on the impact in schools. *Educational Research and Policy Practice*, 14, 119–138. <https://doi.org/10.1007/s10671-014-9171-y>
- 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. <https://doi.org/10.4000/rfp.2296>
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2015). Teacher assessment literacy: a review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251–272. <https://doi.org/10.1007/s11092-015-9233-6>
- DGEEC (2023) – Direção-Geral de Estatísticas da Educação e Ciência [Directorate-General for Education and Science Statistics] through <https://www.dgeec.mec.pt/np4/home>
- Eagly, A. H., & Chaiken, S. (1993). *The Psychology of Attitudes*. Harcourt Brace Jovanovich College Publishers.
- European Commission, EACEA, Eurydice, 2020. *Equity in school education in Europe: Structures, policies and student performance*. Eurydice report. Publications Office of the European Union.

- Fernandes, D. (2009). Educational assessment in Portugal. *Assessment in Education: Principles, Policy & Practice*, 16(2), 227–247. <https://doi.org/10.1080/09695940903076055>
- Fernandes, D. (2021a). Formação contínua de professores em tempos pandémicos: O caso do Projeto MAIA. *Linhas Críticas*. <https://doi.org/10.26512/lc27202139025>
- Fernandes, D. (2021). Para uma fundamentação e melhoria das práticas de avaliação pedagógica no âmbito do Projeto MAIA [To support and improve pedagogical assessment practices within the MAIA Project]. *Ministério da Educação/ Direção-Geral da Educação*. https://afc.dge.mec.pt/sites/default/files/2021-04/TextoApoio1_Para%20uma%20Fundamentação%20e%20Melhoria%20das%20Práticas%20de%20Avaliação%20Pedagógica.pdf
- 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*. <https://doi.org/10.1080/02607476.2020.1799709>
- 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.
- Goldingr, L. (2002). The power of school culture. *Leadership*, 32(2), 32–35.
- Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. John Wiley & Sons.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, L. R. (2006). *Multivariate Data Analysis*. Pearson Education.
- Jimerson, S. R. (2001a). A synthesis of grade retention research: looking backward and moving forward. *Contemporary School Psychology*, 6, 47–59. <https://doi.org/10.1007/BF03340883>
- 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>
- Le Febvre, D., & Richardson, V. (2002). Staff development in early reading intervention programs: the facilitator. *Teaching and Teacher Education*, 18(4), 483–500. [https://doi.org/10.1016/S0742-051X\(02\)00011-2](https://doi.org/10.1016/S0742-051X(02)00011-2)
- Madaus, G. (1993). A national testing system: manna from above? An historical/technological perspective. *Educational Assessment*, 1(1), 9–26. https://doi.org/10.1207/s15326977ea0101_2
- Marôco, J. (2021). *Análise estatística com o SPSS statistics* [Statistical analysis with SPSS statistics] (8th Edition). Report number.
- Martins, G., Gomes, C., Brocardo, J., Pedroso, J., Carrillo, J., Silva, L., & Rodrigues, S. (2017). Students' profile by the end of compulsory schooling. *Ministério da Educação/Direção-Geral da Educação*. https://www.dge.mec.pt/sites/default/files/Curriculo/Projeto_Autonomia_e_Flexibilidade/students_profile_en.pdf
- NAE. (2021). *Educational Assessment in the covid-19 era and beyond*. National Academy of Education, 1–22. <https://naeducation.org/wp-content/uploads/2021/02/Educational-Assessments-in-the-COVID-19-Era-and-Beyond.pdf>
- Nisbet, I., & Shaw, S. (2022). Fair high-stakes assessment in the long shadow of covid-19. *Assessment in Education: Principles, Policy & Practice*, 29(5), 518–533. <https://doi.org/10.1080/0969594X.2022.2067834>
- Decree-Law No. 55/2018, from July the 6th through <https://diariodarepublica.pt/dr/detalhe/decreto-lei/55-2018-115652962>
- 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. (2021). *The state of school education: One year into the COVID pandemic*. OECD Publishing. <https://doi.org/10.1787/201dde84-en>
- OECD. (2023a). Education at a glance 2023: OECD indicators. *OECD Publishing*. <https://doi.org/10.1787/e13bef63-en>
- OECD. (2023b). PISA 2022 results (Volume II): learning during – and from – disruption. *PISA, OECD Publishing, Paris*. <https://doi.org/10.1787/a97db61c-en>
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307–332. <https://doi.org/10.3102/00346543062003307>
- 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. <https://doi.org/10.1080/0969594X.2022.2067123>

- Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: a three-dimensional model. *Teaching and Teacher Education*, 84, 128–138. <https://doi.org/10.1016/j.tate.2019.05.003>
- Popham, W. J. (2011). Assessment literacy overlooked: a teacher educator's confession. *The Teacher Educator*, 46(4), 265–273. <https://doi.org/10.1080/08878730.2011.605048>
- Ribeiro, V. M., Kasmirski, P. R., de Gusmão, J. B., Batista, A. A. G., Janomini, M. A., & Crahay, M. (2018). Crenças de professores sobre reprovação escolar. *Educação Em Revista, Belo Horizonte*, 34, Article 173086.
- Richardson, V., & Placier, P. (2001). Teacher change. *Handbook of Research on Teaching*, 4, 905–947.
- 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. <https://doi.org/10.1080/10627197.2022.2122953>
- 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. https://run.unl.pt/bitstream/10362/89715/1/Tese_Práticas%20e%20Representações%20acerca%20da%20Retenção%20Escolar.pdf
- Santiago, P., Donaldson, G., Looney, A., & Nusche, D. (2012). OECD reviews of evaluation and assessment in education: portugal 2012. *OECD Reviews of Evaluation and Assessment in Education, OECD Publishing, Paris*. <https://doi.org/10.1787/9789264117020-en>
- 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*. <https://doi.org/10.1590/S0104-40362023003103925>
- Santos, N. N., Pipa, J., & Monteiro, V. (2023). Analysing grade retention beliefs within teachers' psycho-pedagogic beliefs system. *Teaching and Teacher Education*, 121, Article 103939. <https://doi.org/10.1016/j.tate.2022.103939>
- Seabra, F., Teixeira, A., Abelha, M., & Aires, L. (2021). Emergency remote teaching and learning in Portugal: preschool to secondary school teachers' perceptions. *Education Sciences*, 11, 349. <https://doi.org/10.3390/educsci11070349>
- Stopher, P. (2012). *Collecting, managing, and assessing data using sample surveys*. Cambridge University Press.
- Sudman, S., & Bradburn, N. M. (1982). Asking questions: A practical guide to questionnaire design. Jossey-Bass.
- Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th Edition). Pearson.
- Tomchin, E. M., & Impara, J. C. (1992). Unraveling teachers' beliefs about grade retention. *American Educational Research Journal*, 29(1), 199–223. <https://doi.org/10.3102/00028312029001199>
- Torres, L. L., & Alves, M. G. (2024). Democratisation and educational inclusion during lockdown times: perceptions of portuguese teachers. *Education Sciences*, 14(2), 197. <https://doi.org/10.3390/educsci14020197>
- Wiese, E., & Nortvedt, G. A. (2023). Teacher assessment literacy in culturally and linguistically diverse classrooms: a norwegian case study. *Teaching and Teacher Education*, 135, Article 104357. <https://doi.org/10.1016/j.tate.2023.104357>
- 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. <https://doi.org/10.1080/0376835X.2024.2311711>
- Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: a reconceptualization. *Teaching and Teacher Education*, 58, 149–162.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.