

The washback effect of external assessments and the COVID-19 pandemic: Perspectives of head teachers in Portugal

Policy Futures in Education
2025, Vol. 0(0) 1–21
© The Author(s) 2025



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/14782103251411715
journals.sagepub.com/home/pfe



Gabriel Cipriano¹  and Susana da Cruz Martins¹

Abstract

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 cancellation 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, we sought to know the point of view of Portuguese head teachers about the impact of this temporary external assessment suspension in schools. Moreover, we wanted 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.

Keywords

COVID-19 pandemic, examinations, external assessment, high-stakes assessments, washback

Introduction

Worldwide, in 2020, the COVID-19 lockdowns led to the sudden implementation of emergency remote teaching across education systems (Hodges et al., 2020), posing significant challenges for

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

Corresponding author:

Gabriel Cipriano, Centro de Investigação e Estudos de Sociologia (CIES-Iscte), Iscte – Instituto Universitário de Lisboa, Iscte-IUL, Edifício Sedas Nunes, Sala 2N17D, Avenida das Forças Armadas, Lisboa 1649-026, Portugal.

Email: gabriel.cipriano@iscte-iul.pt

school leaders (Wharton-Beck et al., 2024). In many countries, this new context of emergency also involved a shift from large-scale external assessments to teacher-built assessments, placing more responsibility on schools, head teachers, and teachers to leverage classroom assessment as the source for information about student growth and achievement (Babbar and Gupta, 2022; Cooper et al., 2022; NAE, 2021; Nisbet and Shaw, 2022; Panadero et al., 2022).

In Portugal, considering the massive disruption experienced by students during this period (e.g. CNE, 2021; Flores and Gago, 2020; Malet Calvo et al., 2022), the Portuguese government also decided to approve extraordinary legislation for the conclusion of the 2019/2020 school year (see Decree-Law No. 14-G/2020). This new legal framework 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 in Portugal; 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, 2014).

It is important to note that assessment of learning, in the Portuguese school system, operates at four key levels: education system, school, teacher, and student. Schools benefit from some autonomy in the organization of the various components of assessment, at the student, teacher, and school levels (Santiago et al., 2012). Moreover, 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 assessment policies to internal contexts and practices. Note too that, worldwide, it is becoming increasingly common for schools to incorporate student test scores in large-scale assessments for accountability purposes (Smith and 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 external assessment of learning suspension during the COVID-19 pandemic, and the model to come in a post-pandemic context.

To contribute to this reflection, we have 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 one, 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 and Cheng, 2017). Consequently, to allow the emergence of these contextual factors during the research, we have decided to implement a qualitative approach to collect the data.

To report our research in this article, in the following section, a literature review was made regarding external assessment of learning purposes and consequences. The subsequent section outlines the methodological framework, making explicit the methods used to collect data, procedures, participants, database characteristics, and data analysis. The next section presents the main results in relation to the research questions formulated earlier, followed by a discussion. Finally, the article concludes by summarising the main findings and implications of the study.

Literature review

Consequences of external assessment tests in education systems

External assessment of learning usually involves the application of a common test, based on prescribed syllabi in the curriculum, with emphasis on written tasks, serving a variety of functions, and it is set and/or controlled by an agency external to the schools from which candidates come, at a state or national level (Kellaghan and Madaus, 2003). External assessment tests, such as examinations, have been employed for a variety of reasons and purposes, for instance, as a means for policymakers to influence instruction and learning in schools, and as a form of accountability (Madaus and Russell, 2010). With a strong ideological orientation shaped by certain representations of schooling, educational stakeholders have, over the years, and alongside the development of school systems worldwide, held different perspectives, conceptions, and assumptions regarding external assessment tests (Machado et al., 2022). Nowadays, in most countries, such 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).

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 and Eggen, 2012). On the one hand, low-stakes assessments aim to influence pedagogical practices toward 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. Yet, the reflexive logic promoted by the external assessment tests' results should be sufficient to produce a process of change (see, e.g. Kim and Choi, 2025). 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; it may involve sanctions and/or financial rewards for teachers depending on achieved 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 and Eggen, 2012).

According to Alderson and Wall (1993), large-scale tests are 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 responsible for putting 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, standardization, and conformity (Madaus and 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 (Kellaghan and Madaus, 2003; Madaus and Russell, 2010; Mons, 2009; Muskin, 2017).

In educational assessment, this phenomenon is referred to as *washback*, which occurs when testing (often large-scale and high-stakes), and, in particular, the uses of test scores and the decisions made on their basis, influence the stakeholders involved and, potentially, their teaching and learning (Tsagari and Cheng, 2017). It is important to note that the washback phenomenon and its negative effects are

not caused by the test itself. Negative impacts do not arise from typical sources of test invalidity, such as construct under-representation or construct-irrelevant variance (Messick, 1996). Instead, test impacts are more directly associated with the ways in which test scores are used, rather than with their interpretation (Gipps, 2011). The stakes attached to test scores often drive teachers, students, and other stakeholders toward behaviours that result in many unintended outcomes (Madaus and Russell, 2010). As Fernandes (2019) further emphasizes, the higher the consequences attached to test scores, the greater the tendency to narrow the curriculum, which, in turn, exacerbates these negative effects.

External assessment tests in the Portuguese education system up to the COVID-19 pandemic

In Portugal, schooling is compulsory for children aged from 6 to 18 years old (Decree-law No. 85/2009). Basic education lasts 9 years divided into 3 cycles of studies. Secondary education lasts 3 years and there is a general path and a professional path that students can choose to finish compulsory education, as shown in Figure 1.

External assessment tests are under the responsibility of the *Instituto de Avaliação Educativa* (IAVE) [National Institute of Educational Assessment] and are compulsory for all students. Until the COVID-19 pandemic (see Figure 1), students took *Provas de Aferição* (PA) in the 2nd, 5th, and 8th grades, which served as national gauging tests without impact on final grades. At the end of basic education (9th grade), students sat for the compulsory *Provas Finais de Ciclo* (PFC) in Portuguese and Mathematics, which accounted for 30% of the final grade in these subjects and certified the completion of basic education. In secondary education, students in the scientific-humanistic track completed four *Exames Nacionais* (EN) in the 11th and 12th grades. These exams certified the completion of secondary education and could also be used for university admission (Decree-Law No. 17/2016; Decree-Law No. 55/2018). Students in the professional track were not required to take the EN and instead completed the *Provas de Aptidão Profissional* (PAP), which are school-based vocational aptitude tests (Decree-Law No. 55/2018).

Consequences of external assessment tests in Portugal

Fernandes (2019) states that PA can be considered very low-stakes assessments; PFC can be considered low-stakes assessments; and EN, in secondary education, can be considered medium-stakes assessments, as they can have a significant weight in accessing higher education. Formally,

Age	6	7	8	9	10	11	12	13	14	15	16	17
Level/Cycle	Basic education									Secondary education		
	1 st cycle				2 nd cycle		3 rd cycle					
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			PA			PA	PFC		EN	EN

Figure 1. External assessment in the Portuguese compulsory school system.

PA – *Provas de Aferição*; PFC – *Provas Finais de Ciclo*; EN – *Exames Nacionais*. Sources: Decree-law No. 85/2009; Decree-law No. 55/2018.

medium or high consequences exist only for students at 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 in policies, it should be noted that a same test can have different stakes depending on the stakeholders' corporative and/or personal goals. Thus, the same test results can have different consequences for educational actors at the same level. For example, for students, the EN carry different stakes depending on their intention to enrol (or not) in higher education, and the admission score required for the specific course they wish to pursue. 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 role in generating school rankings based on students' test results. 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 operationalization. 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.

Considering the effects of external assessment tests and their implications for the functioning of the education system, the Portuguese National Council of Education (CNE) had already noted, prior to the pandemic, that the weight assigned to EN for access to higher education created pressures with negative effects on the objectives of secondary education ([Lourtie, 2020](#)). This recommendation was made in the light of the 'Students profile by the end of compulsory education' (see [PASEO, 2017](#)), which is a humanistic basis profile, focused on the person and human dignity, designed to serve as a guiding framework for all compulsory schooling. But suddenly, the COVID-19 pandemic broke out.

Portuguese external assessment model in the COVID-19 pandemic era

In 2020, to respond to the COVID pandemic emergency, all PA and PFC were cancelled. Moreover, the conclusion of all grade levels in basic and secondary education, including 9th, 11th, and 12th grades, depended only on teacher-constructed assessments. EN at the end of the 11th and 12th grades were still performed, only, for higher education access in subject areas that students choose as admission tests (Decree-Law No. 14-G/2020). In this school year (2019/2020), the absence of PFC and EN to conclude basic and secondary education reinforced a trend towards an increasing percentage of students finishing basic education and secondary education (97.9% and 84.6%, respectively – source: [DGEEC, 2023](#)), and increased the number of students gaining access to higher education in 2020/2021 (133,322 students – source: [DGE, 2023](#)). In fact, all these figures recorded

the highest value ever recorded in Portugal (Cipriano and Martins, 2023). As a result, the solutions found in 2020 seem to have been the subject of a fairly consensual acceptance, and, for that reason, in 2021, 2022 and 2023, unlike some other countries like the UK (see, e.g. Nisbet and Shaw, 2022), the solutions put in place by the Portuguese government remained the same as those of 2020. However, little is yet known of what consequences this temporary suspension had on schools.

Research opportunities and questions

With our research, we wanted to find out 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 the external assessment model should look like in the post-pandemic context. Consequently, two research questions were formulated:

RQ1: Was the cancelation of external assessment of learning during the COVID-19 pandemic a good decision? Why?

RQ2: Should external assessment of learning be reintroduced for the conclusion of basic and secondary education? Why?

Methods

To account for the local educational context, as well as the broader social, political, and economic factors influencing teaching and learning in relation to a test, examination or a testing system as proposed by Tsagari and Cheng (2017), a qualitative research approach was designed, with semi-structured interviews, to be 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 with regard to this topic (Creswell, 2009).

Sampling, procedures, and data analysis

Given the aims of this research to capture local educational contexts, as well as the larger social and administrative factors, a sampling process close to the theoretical sampling process (Bryman, 2012) was implemented, considering different i) School cluster type, ii) School administrative and governance characteristics, and iii) Regions by Nomenclature of Territorial Units II (NUT II). As potential participants for this study, we decided to exclude head teachers from the autonomous regions of Azores and Madeira due to their autonomous administrative characteristics when compared with the Portuguese central government. These two regions have dedicated quotas for higher education access, meaning that the effects of the EN are likely different there. Although examining the effects of EN in these two regions could be interesting for a future study, these schools' head teachers were excluded from the present research. In mainland Portugal, head teachers from private schools were also excluded due to their private administrative characteristics in relation to the Ministry of Education policy. As the mission and objectives of private schools do not necessarily align with those of public schools, research suggests that they face greater pressure regarding access to higher education, with evidence of higher grade inflation in internal assessments compared to public schools (see, e.g. DGEEC, 2025). Although it could be interesting to examine

how pandemic-related cancellations affected the functioning of private schools, the present study focuses exclusively on the public school system; and therefore, private schools were not included in the present research. 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 schools that are just professional schools were also excluded, as for this type of school there are no external assessments tests, relying only on aptitude vocational tests (*Provas de Aptidão Profissional*) performed at a school level to conclude compulsory education.

Following this, 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. Head teachers who voluntarily agreed to participate were interviewed via videoconference from mid-October 2022 until the end of February 2023. As interviews were conducted and recorded, a content analysis was performed ([Bardin, 2014](#)) through an open coding process ([Rädiker and Kuckartz, 2020](#); [Strauss and Corbin, 1990](#)) using MAXQDA 2022 software. To ensure consistency and rigour, all interviews were conducted, transcribed, and coded by the first author under the supervision of the second author. Coding followed an open coding approach ([Strauss and Corbin, 1990](#)), in which data were examined without pre-imposed categories. Throughout the process, the authors held ongoing discussions about emerging categories, collaboratively resolving uncertainties or ambiguities. This approach allowed the identification of local educational contexts as well as broader social, political, and economic factors influencing teaching and learning in relation to the Portuguese large-scale testing system ([Tsagari and Cheng, 2017](#)), while remaining open to new insights and capturing the richness of participants' perspectives.

[Figure 2](#) 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 had been achieved after conducting 32 interviews, with a total of 24 hours and 46 minutes length.

Participants

[Table 1](#) shows the characteristics of the schools from which the interviewed head teachers came. Head teachers from Basic + Secondary school clusters, with regular administration in the Centre Region of Portugal are predominant in the sample.

It should be noted that most schools in Portugal are organized in Basic + Secondary school clusters (from the 1st grade until the 12th grade). In a smaller number, there also exist school clusters with basic education only (from the 1st until the 9th grade) and schools with secondary education only (from the 10th until the 12th grade).

Regarding Administrative and governance characteristics, most public schools in Portugal are regular schools, but some schools have celebrated an autonomy contract with the Ministry of Education for the development of their own educational project (see Decree-Law No. 75/2008). Furthermore, 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 areas ([DGE, 2023](#)).

Geographically, mainland Portugal is organized into five Regions when considering Nomenclature of Territorial Units II (NUT II). North region has the highest number of resident

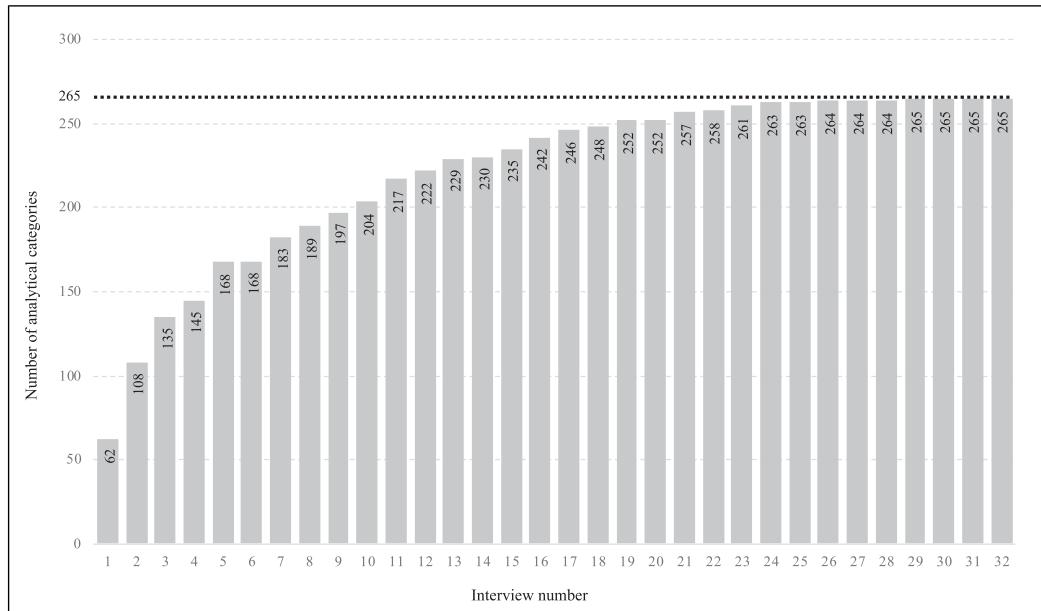


Figure 2. Number of analytical categories by analysed interview number. Source: Primary data, 2023.

population; and Algarve, in the south, has the smallest number of resident population (Source: [INE, 2023](#)).

Results

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, in the discourse of head teachers from schools with just basic education, it was possible to observe a prevalence of segments regarding the need to restructure PA, when compared with other schools' head teachers. In addition, among head teachers from TEIP schools, there was a greater prevalence of segments highlighting views that external assessment devalues students' paths in school and devalues the work that is done by teachers, compared with head teachers from other schools. No relevant content variations were found when considering NUTS II.

Cancelation of external assessment during the COVID-19 pandemic

Regarding RQ1, 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. Public health, physical space in schools, human resources for surveillance, and students'

Table 1. School characteristics by cluster, administrative and governance characteristics, and region.

	Sample		Population	
	n	%	N	%
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

^aIn 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).

emotional well-being conditions were not met to make it viable to perform PA, PFC, and EN with all students. So, cancellation was inevitable and, hence, a good decision:

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

At the time, as much as possible, we had to avoid 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 might have been completely cancelled, but that would have created a major problem for higher education access:

The goal of external assessment 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 might not have 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 to schools.

Reintroduction of external assessment

Regarding the reintroduction of external assessment in the post-pandemic period (RQ2), the previous consensus is lost. Selecting the most developed categories in the content analysis (categories with $n \geq 14$ 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 3](#). The Code Map displays the similarities among categories in a two-dimensional visualization and is based on the tabular representation in the Code Relations Browser. Codes that co-occur more frequently – meaning they are used similarly across the data – are positioned closer together on the map. The positions of the codes are calculated using classical multidimensional scaling, a multivariate statistical method, while hierarchical cluster analysis with an unweighted average linkage is performed to identify groups of related codes. The symbols (circles) on the map represent the frequency of each code, and the colours correspond to the cluster assignments derived from the analysis (Rädiker and Kuckartz, 2020, pp. 94–96).

Below, on the left side of [Figure 3](#), the dark blue area represents head teachers who favour the reintroduction of PFC and EN, as they consider external assessment to be important primarily for accountability purposes. Hence, they tend to reject the existence of PA because, having no stakes for students, it offers them no meaningful consequence. For these head teachers, they view PFC and EN as crucial for maintaining high standards in schools and as a means of guaranteeing 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. As for PA, these head teachers argue that there must be stakes attached 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. (...) [They] are important as they give seriousness to that school grade, which we no longer had with the pandemic model and PFC cancellation. 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 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 results (HT 28).



Figure 3. Head teachers' profiles regarding external assessment reintroduction. Source: Primary data, 2023.

In green, at the top right of Figure 3, are 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. (...) 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. (...) 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 make it possible to manage the curriculum in schools and classrooms, and it promotes a culture of tests instead of a 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 carried out, 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 access to higher education. 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 had 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, (...) without this examination weight to finish secondary education (HT 27).

Below, on the right side of [Figure 3](#), in light blue, are 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 the existence of external assessments is important, and their existence should be mainly for improvement and gauging purposes. Although PA present a number of problems of different orders, they consider that PA should continue to exist, but should 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 take into consideration. 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, in internal assessment practices and 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. EN promote a school ranking based only on EN results, which is unfair to students and schools' accountability. Hence, they consider that the pandemic model could be kept, with EN being performed just to access higher education, as this would be 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. (...) 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).

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 the existence of external assessment is important. 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](#)).

Irrespective of profiles, 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 occur within a school subject, across subject areas, and across grades. This is in line with previous research stating 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 and Russell, 2010](#)); and that the greater the consequences given to test scores, the greater the dysfunctions ([Fernandes, 2019](#)).

Many head teachers, particularly those in the light blue profile of [Figure 3](#), reported that the additional use of EN for higher education admission makes them high-stakes tests. Consequently, EN in secondary education become the driving force behind educational action, rather than PASEO – the students' profile at the end of compulsory school that should be the beacon of the whole education system. This fact is in line with the recommendation made by the National Council of Education ([Lourtie, 2020](#)), stating that EN create a pressure to prepare students as candidates for higher education, with negative effects on the broader objectives of secondary education set out in PASEO.

For many head teachers in the green profile of [Figure 3](#), EN centrality to access higher education and consequent dysfunctions in compulsory education functioning constitute 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. Thus, when it comes to understanding who suffers greatest harm from these negative effects, they say that EN are most harmful to low-achievement students who just want to finish secondary education and cannot because they fail at the EN. In all likelihood, they will be unable to finish secondary education, or they will finish it later, in the recurrent education system (2nd chance education for adults), attending school in the evening (after work), no longer with EN, and this is more expensive for everyone. The COVID-19 pandemic may have offered an opportunity to reconsider the distribution of responsibility between secondary and higher education ([Sá et al., 2021](#)). Consequently, various head

teachers consider that the pandemic model of external assessment could usefully remain, as it was less harmful, especially for low-achievement students.

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](#)); PFC and EN were 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 aims such as controlling, monitoring, certification, and selection, which are not exactly focused on learning improvement ([Kellaghan and Madaus, 2003](#)). Note that if the primary purpose of assessment is to support the learning goals of education, discussions on the relationship between learning and assessment need to be advanced further and positioned at the forefront of high-stakes and large-scale testing. Moreover, theories of learning and assessment for learning should be developed in closer connection. Despite this, assessment and learning theories remain largely separate fields, with significant potential for greater integration ([Baird et al., 2017](#)).

Yet, for head teachers in the dark blue profile of [Figure 3](#), the reintroduction of PFC and EN is viewed as essential for maintaining educational quality and accountability. They emphasize that these tests with stakes are the most reliable mechanism to ensure comparability across schools and to safeguard national standards, both for student achievement and institutional accountability. It should also be noted that the existing PA, PFC, and EN, like 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 and Stobart, 2003](#)). In this framework, PFC and EN are perceived as fostering student commitment and teacher responsibility by attaching tangible consequences to performance, unlike PA, which are regarded as less reliable for not having stakes ([Machado et al., 2022](#)). From these head teachers' perspective, without such stakes, external assessments risk being disregarded by stakeholders, undermining both their validity and reliability as tools for monitoring learning outcomes.

Summoning, in some way, the Stufflebeam 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 test results. This fact was also highlighted by some head teachers from the light blue profile of [Figure 3](#), stating that learning contexts are not considered when designing external assessment tests, not even contexts are considered in the interpretation of test scores. Thus, external assessment tests promote a school ranking based, mainly, on pure external test results ([Torres et al., 2019](#)), which can be highly unfair for TEIP schools and other schools placed in less favourable socio-economic contexts that do 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 assessments ([NAE, 2021](#)); and we must not neglect them for the future.

For some head teachers, namely, the ones in the green profile of [Figure 3](#), 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's \(2019\)](#) considerations, adding that we should evolve towards another generation of external assessments, based on socially constructed principles, focused on student learning and improvement. Possibly, for some head teachers in the green profile of [Figure 3](#), 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 and Stobart, 2003](#)).

At the end of this discussion, beyond the different perspectives and profiles analysed, it is important to emphasize that head teachers broadly agree on the reintroduction of external assessment in the future. The key divergence lies in its intended purposes and format.

Conclusion

The lockdowns caused by the COVID-19 pandemic led to the replacement of many large-scale summative assessments with teacher-designed assessments ([Cooper et al., 2022](#)). Such crisis periods may serve as windows of opportunity to implement reform packages that would be unimaginable during normal times ([Boin and Hart, 2003](#)), and the cancellation of external assessments for the conclusion of basic and secondary education appears to have been one such unprecedented event.

During the pandemic, the cancellation of most external assessments affected the organization of internal assessments in schools. This research shows that the shift in this paradigm, even if temporary, sparked a constructive discussion about the purposes of external assessments. It also demonstrates how external assessments significantly influence the daily management of Portuguese schools ([Machado et al., 2022](#); [Torres et al., 2019](#)) and prompted a questioning of the role of EN in the compulsory education system and their role in students' transition to higher education. Therefore, the changes observed during the pandemic provide a valuable starting point for re-thinking and evaluating the division of responsibilities between secondary and tertiary education ([Sá et al., 2021](#)). Furthermore, from the perspective of many interviewed head teachers, external assessments in Portugal, particularly EN, affect low-achieving and socio-economically disadvantaged students and have negative consequences for the functioning of the education system, while their cancellation had positive effects by mitigating some of these impacts. These experiences highlight the need to critically evaluate the effectiveness and fairness of current higher education access mechanisms, particularly in terms of equity and inclusivity, as high-stakes assessment policies can exacerbate existing inequalities and raise important questions about balancing standardization with social justice in both the completion of secondary education and higher education admissions ([Sá et al., 2021](#)). Such adaptations during the pandemic also illustrate the potential to decouple these responsibilities, enabling secondary education to prioritize broader educational outcomes while encouraging higher education institutions to adopt more flexible and context-sensitive admission mechanisms.

These findings resonate with broader ethical concerns raised by [Buchanan et al. \(2021\)](#) and international reflections on educational transformations prompted by the COVID-19 pandemic ([Jandrić et al., 2022](#)). Both works highlight how the pandemic stimulated educators worldwide to reconsider the foundations of meaningful learning and assessment. [Buchanan et al. \(2021\)](#) argue that educational ethics must address systemic dilemmas such as standardization, accountability, and the marginalization of diverse learners. Their call for a relational and context-sensitive approach to

ethics in education supports the need to rethink large-scale testing models. As this study shows, the stakes attached to EN in Portugal disproportionately affect vulnerable students, raising questions about fairness and the ethical responsibilities of educational institutions. Future assessment reforms should be guided not only by psychometric principles but also by a collective ethical commitment to equity, inclusion, and human dignity, aligning with the PASEO.

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 standardized assessment systems are developed (Mons, 2009). This contextual influence was particularly evident in our research, which revealed a striking consensus among interviewed head teachers regarding the cancellation of external assessments during the COVID-19 pandemic, followed by complete discord afterwards. However, this initial consensus may have arisen because teachers often adjust their positions on assessment when confronted with changing circumstances (Cipriano and Martins, 2025; Xu and Brown, 2016). This became especially clear in the post-pandemic period, when the earlier consensus fractured, giving rise to a landscape of divergent views on the reintroduction of standardized assessments. Such discord highlights that standardized assessment systems are not merely technical instruments but are embedded within complex socio-political and institutional narratives. Thus, the pandemic served as a *wake-up call*, exposing latent tensions within education systems and prompting a critical re-evaluation of the role and legitimacy of external assessments.

It is also important to acknowledge that this study excluded private schools and schools from the autonomous regions of Azores and Madeira. This limitation should be considered when interpreting the previous findings, as the experiences and perspectives of these schools' head teachers may differ from those included in the research. Consequently, the generalizability of the results to all Portuguese schools may be constrained, and future studies should aim to include these contexts to provide a more comprehensive understanding of assessment practices and their implications. Yet, the findings of this study suggest several important implications for policy:

First, policymakers should recognize the significant influence that external assessments exert on the daily functioning of schools and the teaching-learning process. Any reform or reintroduction of standardized assessments, such as EN, must consider their impact on both teachers' practices and students' experiences, particularly for those who are low-achieving or socio-economically disadvantaged.

Second, policymakers should remain attentive to the potential for change, and crisis periods, such as the COVID-19 pandemic, provide valuable insights into the flexibility of high-stakes assessment systems. Evidence from New Zealand, for example, located exactly on the opposite side of the world relative to Portugal, shows that high-stakes assessments operate within dynamically evolving systems, where compromises can be made among competing tensions, national policies, and community expectations. Opportunities for transformation are constrained by existing systems of knowledge, power, and practice. However, where there is collective will and coordinated action, meaningful change can occur (Poskitt, 2022).

Third, the questioning of the role of EN in students' transition to higher education underscores the need for a coordinated approach between secondary and tertiary education systems. Hence, this discussion should take place across both systems, considering the extent to which large-scale assessments can effectively serve multiple purposes, including monitoring the education system, improving learning outcomes, promoting inclusion and equity, supporting the certification function of compulsory schooling in line with PASEO principles, and ensuring that external assessments also fulfil selection functions for higher education.

Finally, it should be highlighted that there is an almost natural predisposition to associate examinations with high standards, rigour, and a means of guaranteeing quality in education systems

(Fernandes, 2014, 2019). Moreover, for many head teachers in Portugal, assessment is still perceived as an instrument for holding stakeholders accountable and exerting 'control' (Santiago et al., 2012). The growing influence of neoliberal and neoconservative agendas in Portugal, along with the public promotion of school excellence based on students' academic performance in the *Exames Nacionais* (EN) (Torres et al., 2019), has created additional pressure to maintain external assessment tests as tools of accountability for teachers, head teachers, and schools. As a result, there has been strong resistance to changing the purposes and formats of external assessments, especially in high-stakes contexts (Cipriano and Martins, 2024). That said, in 2024, within the post-pandemic context and with a new government in charge, the PA were replaced by a new model of computer-based low-stakes assessments with no stakes for students, the PFC were reintroduced without changes to their number or weighting but were moved to a computer-based format, and the EN were reinstated in paper format with minor adjustments to their number and weighting, aimed at reducing unintended consequences (see Decree-Law No. 62/2023). There is evidence that policies aimed at reducing the weighting of exit exams have some merit as a way to mitigate their negative washback on teaching and assessment practices. Nevertheless, there is also evidence suggesting that simply reducing the weighting is insufficient to achieve this goal (Slomp et al., 2020). Future studies in Portugal, beyond 2025, will likely confirm this.

ORCID iD

Gabriel Cipriano  <https://orcid.org/0000-0002-0991-5499>

Ethical considerations

All participants involved in this study provided informed consent prior to their participation. The research procedures complied with ethical standards for research involving human subjects, ensuring confidentiality, voluntary participation, and the right to withdraw at any time without consequences. The study was reviewed and approved by the institutional ethics committee.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: 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 UID/3126/2025.

Acknowledgments

The authors would like to express their sincere gratitude to the 32 head teachers who generously spared their valuable time to participate in interviews for this research project. Their openness, insights, and willingness to engage were essential to the success of the study. The authors also acknowledge their important role in facilitating connections between schools and research, thereby enabling meaningful collaboration between educational practice, academic research, and the development of informed and improved educational policies.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Alderson JC and Wall D (1993) Does washback exist? *Applied Linguistics* 14: 115–129.

Babbar M and Gupta T (2022) Response of educational institutions to COVID-19 pandemic: an inter-country comparison. *Policy Futures in Education* 20(4): 469–491.

Baird JA, Andrich D, Hopfenbeck TN, et al. (2017) Assessment and learning: fields apart? *Assessment in Education: Principles, Policy & Practice* 24(3): 317–350.

Bardin L (2014) *Análise de conteúdo [Content analysis]*. Edições, Vol. 70.

Boin A and Hart PT (2003) Public leadership in times of crisis: mission impossible? *Public Administration Review* 63(5): 544–553.

Bryman A (2012) *Social Research Methods*. 4th edition. Oxford University Press.

Buchanan RA, Forster DJ, Douglas S, et al. (2021) Philosophy of education in a new key: exploring new ways of teaching and doing ethics in education in the 21st century. *Educational Philosophy and Theory* 54: 1178–1197.

Cipriano G and Martins SC (2021) Beliefs on assessment and grade repetition among teachers in Portugal. *Revista Meta: Avaliação* 13(39): 248–273.

Cipriano G and Martins Sda C (2023). Avaliação das aprendizagens em Portugal durante a pandemia. In Fonseca D, Neto-Mendes A, Gonçalves M, et al. (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 and Martins Sda C (2024) Introducing technologies into national large-scale testing: are we ready? *Education Policy Analysis Archives* 32(20): 1–21.

Cipriano G and Martins Sda C (2025) Examining the effects of the COVID-19 pandemic and professional development on teachers' grade repetition beliefs and practice. *Educational Research for Policy and Practice*.

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.

Cooper A, Deluca C, Holden M, et al. (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>

Creswell JW (2009) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 3rd edition. Sage Publications.

DGEEC – Direção-Geral de Estatísticas da Educação e Ciência (2025) Análise das classificações internas nos cursos científico-humanísticos em estabelecimentos públicos e privados de Portugal Continental, 2018/19 – 2023/24 [Analysis of internal grades in scientific-humanistic courses in public and private schools in mainland Portugal, 2018/19 – 2023/24]. <https://www.dgeec.medu.pt/api/ficheiros/67b5c449845b70917a3c7690>

Fernandes D (2014) Avaliações externas e melhoria das aprendizagens dos alunos: questões críticas de uma relação (im)possível [External assessment and improvement of students learning: critical issues of an (im) possible relationship]. In: Ferreira A (ed) *Avaliação Externa e Qualidade das Aprendizagens*. Conselho Nacional de Educação, 21–49.

Fernandes D (2019) Avaliações externas e aprendizagens dos alunos: uma reflexão crítica [External assessment and students learning: a critical reflection]. *Linhas Críticas*.

Flores MA and Gago M (2020) Teacher education in times of COVID-19 pandemic in Portugal: National, institutional and pedagogical responses. *Journal of Education for Teaching* 46: 1–10.

Gipps C (2011) *Beyond Testing (Classic Edition): Towards a Theory of Educational Assessment*. Routledge.

Gipps C and Stobart G (2003) Alternative assessment. In: Stufflebeam D, Kellaghan T and Wingate L (eds) *International Handbook of Educational Evaluation*. Kluwer, 549–575.

Hodges C, Moore S, Lockee B, et al. (2020) The difference between emergency remote teaching and online learning. *EDUCAUSE Review*. <https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remoteteaching-and-online-learning>

Jandrié P, Fuentes-Martinez A, Reitz C, et al. (2022) Teaching in the age of Covid-19—The new normal. *Postdigital Science and Education* 4(4): 877–1015.

Kellaghan T and Madaus G (2003) External (public) examinations. In: Stufflebeam D, Kellaghan T and Wingate L (eds) *International Handbook of Educational Evaluation*. Kluwer, 577–602.

Kim Y and Choi T-H (2025) The influence of the programme for international student assessment on educational governance situated in the institutional setting of South Korea. *Policy Futures in Education* 23(1): 31–47.

Lourtie P (2020) *O acesso ao ensino superior e a articulação com o ensino secundário: Recomendação [Access to higher education and its articulation with secondary education: Recommendation]*. Conselho Nacional de Educação.

Machado EA, Flores MA, Pereira D, et al. (2022) Políticas e práticas de avaliação externa dos alunos na perspetiva dos diretores: o caso das provas de aferição [Policies and practices of external assessment of students from the perspective of principals: the case of low stakes assessment tests]. *Revista Portuguesa de Investigação Educacional* 23: 1–26.

Madaus G and Russell M (2010) Paradoxes of high-stakes testing. *Journal of Education* 190: 21–30.

Malet Calvo D, Cairns D, França T, et al. (2022) There was no freedom to leave': Global South international students in Portugal during the COVID-19 pandemic. *Policy Futures in Education* 20(4): 382–401.

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 [Theoretical and real effects of standardized evaluation policies]. *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. <https://repositorio.minedu.gob.pe/handle/MINEDU/5525>

NAE (2021) Educational assessments 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 and 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>

Panadero E, Fraile J, Pinedo L, et al. (2022) Changes in classroom assessment practices during emergency remote teaching due to COVID-19. *Assessment in Education: Principles, Policy & Practice* 29(3): 361–382.

PASEO (2017) *Perfil dos alunos à saída da escolaridade obrigatória [Student's profile by the end of compulsory schooling]*. Ministério da Educação / Direção-Geral da Educação.

Poskitt J (2022) COVID-19 impact on high stakes assessment: a New Zealand journey of collaborative adaptation amidst disruption. *Assessment in Education: Principles, Policy & Practice* 29(5): 575–595.

Rädiker S and Kuckartz U (2020) *Focused Analysis of Qualitative Interviews with MAXQDA: Step by Step*. Maxqda Press.

Rozenwajn E and Dumay X (2014) *Les effets de l'évaluation externe sur les pratiques enseignantes: une revue de la littérature [The effects of external assessment on teaching practices: a review of the literature]*. Revue française de pédagogie.

Sá C, Sin C, Pereira F, et al. (2021) *Estudantes nacionais e internacionais no acesso ao ensino superior: Quem são, que escolhas fazem e como accedem ao mercado de trabalho [National and international students in access to higher education: Who they are, what choices they make, and how they enter the labor market]*. EDULOG — Belmiro de Azevedo Foundation. <https://www.edulog.pt/storage/app/uploads/public/632/454/79e/63245479ef2ab111642163.pdf>

Santiago P, Donaldson G, Looney AE, et al. (2012) *OECD Reviews of Evaluation and Assessment in Education: Portugal 2012*. OECD Publishing.

Slom D, Marynowski R, Holec V, et al. (2020) Consequences and outcomes of policies governing medium-stakes large-scale exit exams. *Educational Assessment, Evaluation and Accountability* 32: 431–460.

Smith WC and Holloway J (2020) School testing culture and teacher satisfaction. *Educational Assessment, Evaluation and Accountability* 32: 461–479.

Stobart G and Eggen T (2012) High-stakes testing - value, fairness and consequences. Assessment in education: principles. *Assessment in Education: Principles, Policy & Practice* 19(1): 1–6.

Strauss A and Corbin J (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Sage publications.

Stufflebeam D (2003) The CIPP model for evaluation. In: Stufflebeam D, Kellaghan T and Wingate L (eds) *International Handbook of Educational Evaluation*. Kluwer, 31–62.

Torres LL, Palhares JA and Afonso AJ (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 and Cheng L (2017) Washback, impact, and consequences revisited. In: Shohamy E, Or I and May S (eds) *Language Testing and Assessment. Encyclopedia of Language and Education*. Springer.

Verdasca J (2013) Rankings escolares: «a César o que é de César e a Deus o que é de Deus» [School rankings: “to Caesar what is from Caesar and to God what is from God”]. *Educação: Temas e Problemas* 12-13: 294–327.

Wharton-Beck A, Chou CC, Gilbert C, et al. (2024) K-12 school leadership perspectives from the COVID-19 pandemic. *Policy Futures in Education* 22(1): 21–42.

Xu Y and Brown GT (2016) Teacher assessment literacy in practice: a reconceptualization. *Teaching and Teacher Education* 58: 149–162.

Legislation

Decree-Law No. 75/2008.

Decree-Law No. 85/2009.

Decree-Law No. 17/2016.

Decree-Law No. 54/2018.

Decree-Law No. 55/2018.

Decree-Law No. 14-G/2020.

Decree-Law No. 22-D/2021.

Decree-Law No. 27-B/2022.

Decree-Law No. 22/2023.

Decree-Law No. 62/2023.

Other sources

DGAE (2019) – Direção-Geral da Administração Escolar. [Directorate-General for School Administration]. https://www.dgae.medu.pt/download/recrutamento/outras/201819/20190307_Rec_Codigos-AE_ENA-2019_2020.pdf (last accessed in July 2023).

DGE (2023) – Direção-Geral da Educação. [Directorate-General for Education]. <https://www.dge.mec.pt> (last accessed in July 2023).

DGEEC (2023) – Direção-Geral de Estatísticas da Educação e Ciência [Directorate-General for Education and Science Statistics]. <https://www.dgeec.medu.pt> (last accessed in July 2023).

GesEdu (2023) – Instituto de Gestão Financeira da Educação [Institute for Financial Management of Education]. <https://www.gesedu.pt/PesquisaRede> (last accessd in July 2023).

INE (2023) Instituto Nacional de Estatística [National Institute of Statistics]. <https://www.ine.pt> (last accessed in July 2023).

Author biographies

Gabriel Cipriano is an Invited Professor in the Department of Social Research Methods at the School of Sociology and Public Policy of Iscte-IUL, and a researcher at CIES-Iscte (Centre for Research and Studies in Sociology). With a PhD in Policies of School Administration and Management from Iscte, his main research interests focus on evaluation theories, assessment of learning, evaluation of public policies in education, school administration policies, and school commuting policies.

Susana da Cruz Martins is a sociologist with Ph.D. in Sociology. Associate Professor at the Department of Political Science and Public Policy of Iscte-IUL, she is also a researcher at CIES-Iscte, and member of the Inequality Observatory (CIES-Iscte). Her main research interests are higher education, comparative education systems, education and social mobility, and social inequalities.