education policy analysis archives

A peer-reviewed, independent,
open access, multilingual journal

Arizona State University

Volume 32 Number 20 April 2, 2024 ISSN 1068-2341

Introducing Technologies into National Large-Scale Testing: Are We Ready?[1]

Gabriel Cipriano
&
Susana da Cruz Martins
Centro de Investigação e Estudos de Sociologia (CIES-Iscte),
Iscte – Instituto Universitário de Lisboa
Portugal

Citation: Cipriano, G., & Martins, S. C. da (2024). Introducing technologies into national large-scale testing: Are we ready? *Education Policy Analysis Archives, 32*(20).
https://doi.org/10.14507/epaa.32.8538

Abstract: In recent decades, the potential benefits of introducing technologies into large-scale tests have been much discussed. Yet the path to effective technology use on large-scale testing has fallen short of expectations, especially when these tests have medium or high stakes for students. After a temporary cancelation of external assessment of learning due to the COVID-19 pandemic, the Portuguese government is taking a step forward. With a gradual implementation until 2025, the reintroduction of external assessment of learning in the Portuguese education system foresees all national assessment tests and all national examinations in digital format with the implementation of the *External Assessment Dematerialization Project* (DAVE). However, assessment reforms, such as DAVE, raise concerns within school communities. To identify and analyze these concerns, we conducted 32 semi-structured

[1] 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.

Journal website: http://epaa.asu.edu/ojs/
Facebook: /EPAAA
Twitter: @epaa_aape

Manuscript received: 22/1/2024
Revisions received: 8/3/2024
Accepted: 11/3/2024

interviews with head teachers from mainland Portugal. To understand to what extent these concerns were considered and covered in DAVE's design and implementation, a supplementary interview was conducted with the president of the Portuguese institute responsible for DAVE. In addition, a survey with 2 673 teachers was carried out to find out their degree of agreement with the implementation of DAVE. Results show that DAVE raises many concerns among head teachers, and it is not fully accepted by the teachers.
Keywords: computer-based testing; large-scale testing; digital external assessment; digital high-stakes testing

Introduciendo tecnologías en las pruebas nacionales a gran escala: ¿Estamos preparados?
Resumen: En las últimas décadas, se han discutido mucho los beneficios potenciales de introducir tecnologías en pruebas a gran escala. Pero el camino hacia el uso efectivo de la tecnología en las pruebas a gran escala no ha estado a la altura de las expectativas, especialmente cuando las pruebas tienen implicaciones asociadas para los estudiantes. Después de una cancelación temporal de la evaluación externa de los aprendizajes debido a la pandemia de COVID-19, el gobierno portugués está dando un paso adelante. Con una implementación gradual hasta 2025, la reintroducción de la evaluación externa de los aprendizajes en el sistema educativo portugués prevé todas las pruebas de evaluación y todos los exámenes nacionales en formato digital con la implementación del *Proyecto de Desmaterialización de la Evaluación Externa* (DAVE). Sin embargo, las reformas de las evaluaciones, como DAVE, plantean preocupaciones en las comunidades escolares. Para identificar y analizar estas preocupaciones, se realizaron 32 entrevistas semiestructuradas con directores de escuelas en Portugal continental. Para comprender en qué medida se consideraron y abordaron estas preocupaciones en el diseño y la implementación de DAVE, se realizó una entrevista complementaria con el presidente del Instituto responsable de DAVE. Además, se realizó una encuesta a 2 673 docentes para conocer su grado de acuerdo con la implementación del DAVE. Los resultados muestran que DAVE plantea muchas preocupaciones entre los directores de escuelas y no es totalmente aceptado por los docentes.
Palabras-clave: pruebas por ordenador; pruebas a gran escala; evaluación externa digital; pruebas digitales de alto impacto

Introduzindo tecnologias em testes nacionais em larga escala: Estamos prontos?
Resumo: Nas últimas décadas, os potenciais benefícios da introdução de tecnologias em testes em larga escala têm sido muito discutidos. No entanto, o caminho para um uso efetivo de tecnologias em testes em larga escala ficou aquém das expectativas, especialmente quando os testes têm implicações associadas para os alunos. Depois de um cancelamento temporário da avaliação externa das aprendizagens devido à pandemia de COVID-19, o Governo português está a dar um passo em frente. Com uma implementação gradual até 2025, a reintrodução da avaliação externa das aprendizagens no sistema educativo português prevê todas as provas de aferição e todos os exames nacionais em formato digital com a implementação do *Projeto para a Desmaterialização da Avaliação Externa* (DAVE). No entanto, reformas na avaliação das aprendizagens, como o DAVE, suscitam preocupações nas comunidades escolares. Para identificar e analisar estas preocupações, foram realizadas 32 entrevistas semiestruturadas com diretores de escola de Portugal continental. Para perceber em que medida estas preocupações foram consideradas e abordadas na conceção e implementação do DAVE, foi realizada uma entrevista suplementar com o presidente do Instituto responsável pelo DAVE. Além disso, foi realizado um inquérito por questionário a 2 673 professores para averiguar o seu grau de

concordância com a implementação do DAVE. Os resultados mostram que o DAVE levanta muitas preocupações entre os diretores escolares e não é totalmente aceite pelos professores.
Palavras-chave: testes por computador; testes em larga escala; avaliação externa digital; testes digitais de alto impacto

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

Large-Scale Testing

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

Large-scale testing usually involves the application of a common paper-based test, constructed on prescribed syllabi in curriculum, with emphasis on written tasks, serving several functions, and it is set or controlled by an agency external to the schools from which candidates come, at a state or national level (Kellagham & Madaus, 2003; OECD, 2023). Large-scale tests also have a standardized universal design, to maximize accessibility for all intended test takers, knowing that all test takers should have an unobstructed opportunity to demonstrate their standing on the construct being measured. Nevertheless, when the test is not appropriate for all test takers in the intended population, accommodations can be made, retaining the comparability of test scores. To allow comparability of test scores in large-scale testing, the processes by which test takers' responses are evaluated and scored are also standardized (AERA et al., 2014).

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

Introducing Technologies into National Large-Scale Testing

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

More than 20 years after the millennial beginning, in most countries in the world, large-scale testing continues to play a significant role in education systems (Fernandes, 2019), with an increasing number of countries performing national or central assessments tests with no stakes for students, reflecting the existing trend towards monitoring standards and collecting diagnostic information to support achievement (OECD, 2023). Nevertheless, while sophisticated digital learning platforms, multimedia technologies and wireless communication are transforming what, when and how learning can take place, it seems that transformation in assessment thinking and practice trails behind. In many settings paper-based testing is still seen as the most reliable way to assess educational achievements (Richardson & Clesham, 2021).

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

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

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

Those who advocate the use of computer-based large-scale testing suggest that it might induce positive impacts in education systems and classroom practices. The potential benefits include: a) time, resources and costs savings in the test administration and scoring; b) great score precision with the possibility to monitor teachers' scoring; c) improvements in assessment validity and reliability; d) improvements in test security; e) greater opportunity to evaluate progress over time and to use test results to influence instruction; f) possibilities for new kinds of questions, using multimedia, simulations and other resources to assess sophisticated learning goals; and g) adaptive assessments based on the questions asked and on students' previous responses (IAVE, 2022; Kerrey & Isakson, 2000; Oldfield et al., 2012). Nevertheless, they also recognize the barriers that a shift from a paper-based testing system to a computer-based testing system can bring, such as: a) concerns about the capability of the technology to assess all subjects and its consequences on the reliability and validity of high-stakes assessments; b) comparability; c) lack of staff, time and training to use computer-based testing on a pedagogical perspective; d) students inequal access to practice on the relevant software or devices; e) cost of investment; f) user verification and security issues; g) lack of suitable physical spaces and devices in schools; h) ineffective and/or unreliable broadband, wi-fi and network capabilities at schools;

i) accessibility, namely for students with special educational needs and students with disabilities; and j) the consequences on public opinion about the security, fairness and/or malpractice (Ofqual, 2020; Oldfield et al., 2012).

Considering the barriers and negative impacts that computer-based testing can have, it seems that many governments are being cautious when introducing computer-based technology on large-scale testing, especially when they have associated stakes, as examinations have. While the number of OECD countries using computer-based technology in national/central assessment tests has increased from 8, in 2015, to 21 in 2023, computer-based examination is not yet widely implemented. Only eight countries use computer-based uniform technology for at least one exam (OECD, 2023). These figures might reflect the existing fears and uncertainty about the positive and negative impacts of computer-based examinations on students' pathways. In addition, when implementing a computer-based testing initiative, Tomas et al. (2015) highlight that research on e-assessment has been dominated by a focus on investigating benefits of use rather than building an understanding of development and implementation. When studying the development and implementation of a computer-based testing initiative, their research showed that subtle interplay exists between assessment stakes, type, stages, and modes that should be considered. Therefore, it seems to be crucial to understand the positive and negative impacts of large-scale computer-based tests, and they should be studied and documented in the development and implementation of computer-based testing initiatives (Johnson & Shaw, 2018).

Large-Scale Testing in the Portuguese Education System

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 three study cycles. Secondary education lasts 3 years and there is a general pathway and a professional pathway that students can choose to finish compulsory education.

Large-scale tests in the Portuguese education system are a responsibility of *Instituto de Avaliação Educativa* (IAVE; National Institute of Educational Assessment] and tests are compulsory for all students. In the middle of each cycle of basic education, in the second, fifth, and eighth grades, there are *Provas de Aferição* (PA). PA are national assessment tests that do not have weight on students' grading. Some of these PA are practical tests with standardized procedures, such as PA in Music Education, Sports, and Arts; while other PA are written tests, such as in History and Natural Sciences. At the end of basic education, in the ninth grade, there are *Provas Finais de Ciclo* (PFC) on Portuguese and Mathematics subject areas. PFC are written tests, compulsory for all students, and they aim to evaluate students' performance in Portuguese and Mathematics, certifying the conclusion of basic education, with a weight of 30% in the final classification of these two subject areas. At the end of the 11th (two exams) and 12th grades (two exams), there are compulsory *Exames Nacionais* (EN). EN are written tests that aim to evaluate students' performance and certify the conclusion of secondary education, with a weight of 30% on final classification of subject areas with external assessment test. In addition, these tests' results can also be considered for access to higher education. The conclusion of secondary education on the professional pathway does not imply performing EN. There are, in replacement, *Provas de Aptidão Profissional*, which are aptitude vocational tests set by each school (Decree-law No. 17/2016; Decree-law No. 55/2018). The weight and the number of EN to be performed in secondary education will have small changes from 2024 onwards (See Decree-Law No. 62/2023).

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

As referred before, barriers to implement a computer-based testing initiative may include costs of investment to suppress the lack of suitable physical spaces and devices in schools, ineffective and/or unreliable broadband, wi-fi, and network capabilities, as well as lack of training to use computer-based testing on a pedagogical perspective (Ofqual, 2020; Oldfield et al., 2012). These are costs that many countries cannot afford, including Portugal, without additional lines of investment in their annual educational budget.

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

Within this TD-C20-i01 financial package from the PRR, the *External Assessment Dematerialization Project* (DAVE) is funded and implemented in Portugal (IAVE, 2022; SGEC, 2022). DAVE aims to integrate and dematerialize all the procedures inherent to an external assessment of learning process, from the organizational and logistical point of view until the processes of test preparation, application, and marking. To implement DAVE with the greatest security and contribution of all actors involved in the process, DAVE has been tested by IAVE since 2018, and it is currently being gradually developed until its full implementation in 2025 (Eurydice, 2023; IAVE, 2022).

Research Opportunities and Questions

As proposed by Johnson & Shaw (2018) to study positive and negative impacts on the development and implementation of a computer-based testing initiative, we wanted to understand the perspectives of different stakeholders regarding the implementation of a computer-based testing system in Portugal. Consequently, this paper aims to contribute to the deepening of understanding about the development and implementation of a computer-based large-scale testing system, using Portugal as a case study. Such a study will provide a comprehensive overview of the use of technologies in national large-scale testing, as well as different controversies about this subject. We believe that such a study is very relevant for Portuguese policymakers and broad stakeholders, as well as for other countries facing similar ongoing processes at present or in the near future. We formulated three research questions (RQ), where RQ1 was addressed to schools' head teachers, RQ2 was addressed to the president of IAVE, and RQ3 was addressed to teachers:

RQ1: Are schools ready for the implementation of DAVE? Why?
RQ2: Were head teachers concerns covered in DAVE's design? How?
RQ3. To what extent do Portuguese teachers agree with DAVE?

Methods

To capture positive and negative impacts during the development and implementation of a computer-based testing initiative (Johnson & Shaw, 2018), we documented different moments of DAVE's implementation. In addition, to capture the local educational context, as well as the larger social, political, and economic factors that govern teaching and learning in relation to this new large-scale testing system (Tsagari & Cheng, 2017), a mixed method approach was designed with different stakeholders.[2] In the first part of this study, semi-structured interviews were conducted by videoconference calls with head teachers from different types of public school in different regions of Portugal, during the 2022/2023 school year, before the implementation of PA in digital format in May 2023. These interviews were intended to elicit views and opinions from the participants (Creswell, 2009) regarding this computer-based testing initiative, as well as their concerns about its implications.

In the second part of this study, considering the identified head teachers' concerns about DAVE, a semi-structured interview script was designed, with 11 open-ended questions, addressed to the president of IAVE. The IAVE president's interview was conducted in person, in July 2023, after the implementation of PA in digital format during the 2022/2023 school year, aiming to know how head teachers' concerns were covered in DAVE's design and implementation.

In the third part of this study, to understand to what extent Portuguese teachers agree with the implementation of DAVE, a survey was disseminated in late September 2023 until the end of October 2023, among public and private school communities, before the universal implementation of the ninth-grade PCF in digital format, to be performed during the 2023/2024 school year.

Sampling, Procedures, and Data Analysis

Most schools in Portugal are organised in Basic + Secondary school clusters (Grades 1–12). In a smaller number, school clusters with basic education only (Grades 1–9) exist, as do schools with secondary education only (Grades 10–12). Concerning administrative and governance characteristics, most schools in Portugal have a regular administration directly dependent of the Ministry of Education, but some schools have an autonomy contract with the Ministry of Education for the development of their own educational project (see Decree-Law No. 75/2008). Moreover, the creation of *Territórios Educativos de Intervenção Prioritária* (TEIP) [Educational Territories of Priority Intervention] is a governmental initiative, to better support schools in economically and socially disadvantaged territories (Normative Order No. 20/2012). In a new program phase (TEIP4), school clusters apply for support through an action plan in partnership with the municipalities. Geographically, Portugal is organized into seven regions when considering Nomenclatures of Territorial Units II (NUT II). Two of these regions are the autonomous regions of Azores and Madeira, the other five regions are located in mainland Portugal. Given the purposes of the first part of this research to capture local educational contexts, as well as the larger social and administrative factors before the implementation of DAVE, a sampling process was implemented considering different: i) School cluster type, ii) Administrative and governance characteristics, and iii) Regions by Nomenclature of Territorial Units II (NUT II) in mainland Portugal.

To gather participants for the interviews, an e-mail was sent to head teachers from public schools and school clusters in mainland Portugal promoting this study. As head teachers responded to the sent e-mail to voluntarily participate in this study, interviews were conducted via a videoconference call, from mid-October 2022 until the end of February 2023. As the interviews

[2] This research was approved by an institutional ethics committee, and all participants were provided appropriate informed consent.

were carried out and recorded, a content analysis was performed through an open coding process (Rädiker & Kuckartz, 2020; Strauss & Corbin, 1990) using the MAXQDA 2022 software. No new or relevant data emerged regarding analytical categories in the latest analyzed interviews, and existing categories were already well developed in terms of their properties and dimensions demonstrating variation with well-established relationships among the categories (Bryman, 2012, p. 241). Thus, the research team determined that theoretical saturation was reached after conducting 32 interviews.

In the second part of this research, IAVE president's interview was recorded, and a content analysis was also performed through an open coding process (Rädiker & Kuckartz, 2020; Strauss & Corbin, 1990) using MAXQDA 2022 software. The collected data was then analyzed considering the identified head teachers' concerns.

In the third part of this research, teachers' survey was conducted online, preceded by a pre-test. For the survey dissemination, an e-mail was sent to all head teachers at all public schools in Portugal, including the autonomous regions of Azores and Madeira. It was also sent an e-mail to all private schools that performed EN in the 2022/2023 school year. The emails were sent during the last week of September 2023. The survey was then forwarded by the school's head teachers to all the other teachers in their schools to voluntarily participate on the study, and it was online during the whole month of October 2023. All answers are anonymous, and it is not possible to relate a teacher to another teacher or school. We used descriptive statistics to analyze the survey data.

Participants

Interviews with Head Teachers

Table 1

School Characteristics by Cluster Type, 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

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

Table 1 shows the characteristics of the schools from the 32 interviewed head teachers. Head teachers from Basic + Secondary school clusters, with regular administration in the Centre Region of Portugal are predominant in the sample.

Teachers' Survey

Table 2 shows the characteristics of the 2 673 teachers that participated on the survey, by sex, age, school sector and region. As the sample was formed through a non-probabilistic process, it is possible to observe a sampling bias regarding Sex (χ^2 (1, N = 2 619) = 11.30, $p < 0.001$), Age (χ^2 (3, N = 2 673) = 51.23, $p < 0.001$), School sector (χ^2 (1, N = 2 609) = 28.54, $p < 0.001$) and region (χ^2 (6, N = 2 644) = 622.59, $p < 0.001$). Nevertheless, this is a very large sample and, in general, the proportions (in %) are very similar between the sample and the population.

Table 2

Teachers from the Compulsory School System by Sex, Age, School Sector, and Region

	Sample		Population	
	n	Valid %	*N*	%
Sex [a]				
Female	2 058	78.6	93 999	75.8
Male	561	21.4	30 067	24.2
Age [a]				
< 30 years old	41	1.5	2 640	2.1
Between 30 – 39 years old	119	4.5	9 424	7.6
Between 40 – 49 years old	923	34.5	44 140	35.6
≥ 50 years old	1 590	59.5	67 862	54.7
School sector [a]				
Public	2 439 [b]	93.6 [b]	112 160	90.4
Private	170 [b]	8.8 [b]	11 906	9.6
Region by Nomenclature of Territorial Units II [c]				
North	590	22.1	44 724	33.4
Centre	579	21.7	27 108	20.2
Lisbon Metropolitan Area	733	27.5	36 366	27.2
Alentejo	113	4.2	9 301	6.9
Algarve	147	5.5	6 567	4.9
Autonomous Region of Azores	266	10.0	4 699	3.5
Autonomous Region of Madeira	236	8.9	5 200	3.9

Note: a) The population data refers to the 2021/2022 school year in mainland Portugal, excluding Madeira and Azores, as no statistics were available for these two autonomous regions regarding sex, age, and school sector. b) 64 teachers (2.4%) reported to teach both in public and private sector. c) The population data refer to the 2021/2022 school year, including Madeira and Azores. Sources: DGEEC, 2023; DRAEM, 2023; DREAEA, 2023; Primary data, 2023 (Own calculations).

Results

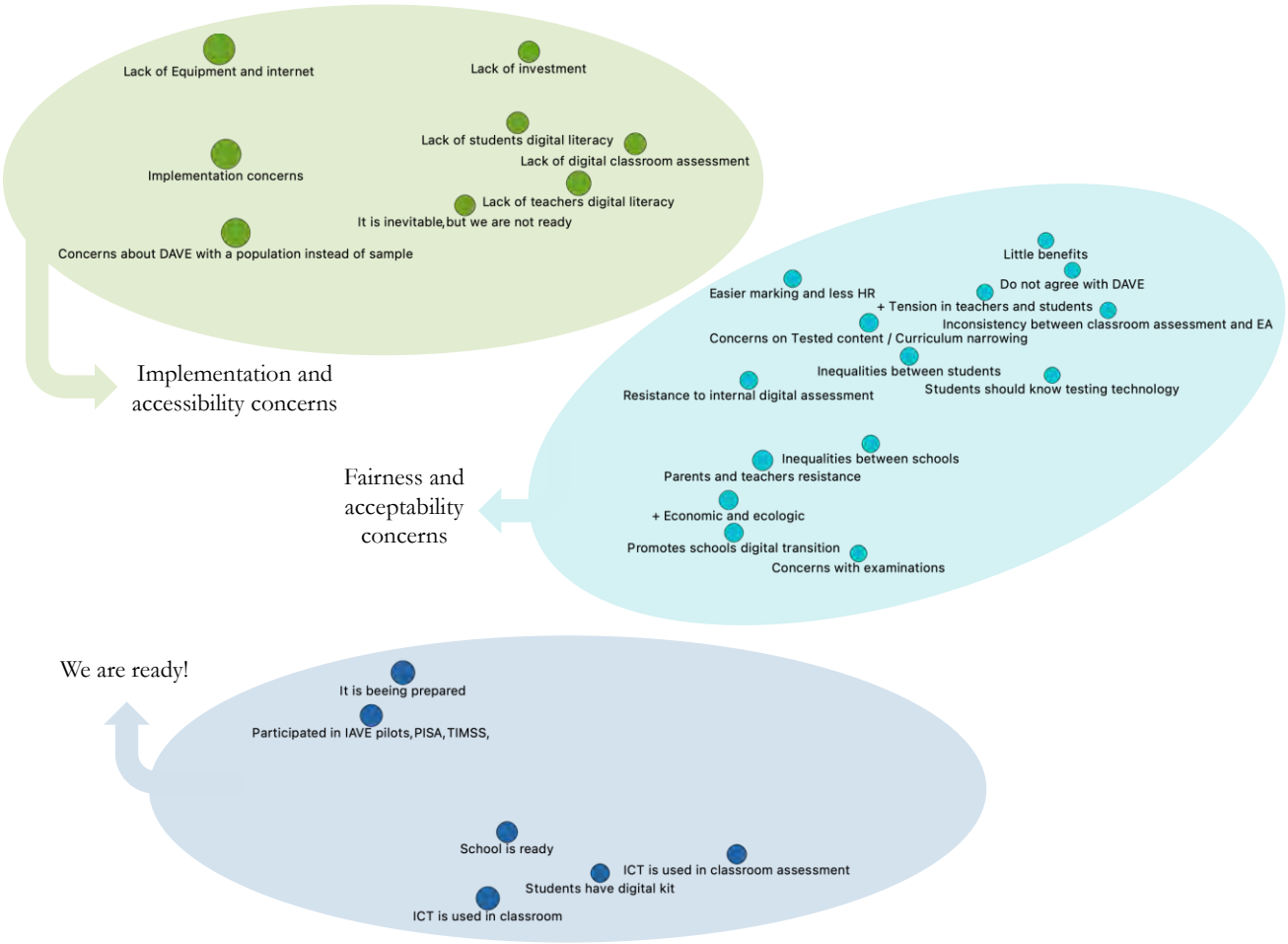
RQ1. Head Teachers' Concerns regarding the Implementation of DAVE

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

Then, performing a relation of codes mapping in MAXQDA, by co-occurrence of codes in the documents, we identified three head teachers' profiles shown in Figure 1. The Code Map displays the similarities of categories in a two-dimensional visualization. The more co-occurrences two codes have, that is, the more similarly they are used in the data, the closer they are placed together on the Code Map. The symbols (circles) sizes used in the Code Map represent the code frequencies. The colors of the codes correspond to the calculated cluster assignments on the map (Rädiker & Kuckartz, 2020, pp. 94-96).

Figure 1

Head Teachers' Concerns and Profiles regarding DAVE



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

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

> At the moment, I am talking to you at this conference-call, and I am not using the school's internet connection provided by the Ministry of Education. I am talking to you through a private connection. Otherwise, we would not be talking. (HT16)
>
> (...) Do schools support an exam under these conditions? For example, I have 150 students who will take the Portuguese exam. But I don't have 150 computers. The Ministry of Education might say that all students have their own computer. But will all computers work at that time? (HT1)
>
> For the introduction of digital approaches in schools and for the maintenance of the equipment, we need to have a technical team in each school, entirely dedicated to these issues. It cannot be the goodwill of teachers to do this. This is a very voluntary way of dealing with things and it is not professional. (HT26)

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

> In schools, we do not have internal tests with these characteristics. This requires a very specific logistics, and it requires that all schools have adequate resources. (HT15)
>
> (...) a lot of work needs to be done with the teachers. Because students cannot use the computer only at the examination. There must be training and a consolidation of practices that must happen before in the classroom. (HT12)
>
> In younger age groups, particularly in the second grade, it is complicated. We know that our children know how to use tablets and mobile phones, but when they get to a computer, it is not quite the same. (HT8)
>
> We know that PISA is done like this, but it is done with a sample. With a sample, it works. With a school population with big differences between them, at the same time, with the internet working as it works in schools... I don't know.... (HT 25)

The dark blue in Figure 2, indicates head teachers who considered that *Schools are ready!* They consider that there is already a regular use of ICT in the classroom, students have their own technological kit with a personal computer and a personal internet access, and there is already some use of computer-based tests in classrooms. Also, most of these schools have participated in other computer-based large-scale tests such as TIMSS, PISA, and PIRLS, as well as in IAVE's pilots. Therefore, for these head teachers, schools are ready:

> We have a particularity that does not occur in many schools: In addition to the [technological] kit that each student received, all students have access to a

> computer at school. But I know that this is a school that does not exist in many places. (HT 31)
>
> At the ninth-grade level, there are already many teachers who do some kind of assessment through digital resources and digital applications. (HT14)
>
> We, by chance, are a school that a few years ago did the Mathematics PA in digital format. IAVE asked us if we wanted to participate in their pilot, we said yes, and it went beautifully. (HT9)

In light blue, at the top right of Figure 2, illustrates how there are head teachers whose major concerns rely on the *fairness and acceptability* of digital assessments. They recognize that computer-based large-scale testing might promote the digital transition of schools, with benefits on economic and ecological issues, involving fewer human resources for the application and marking processes. However, for these head teachers, the benefits are few considering what is at stake, especially in PFC and EN with implications on the students' pathways:

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

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

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

RQ2. Concerns Covered in DAVE's Design and Implementation

Regarding the head teachers' concerns identified in green on Figure 2, IAVE's president recognize that schools have strong restrictions on internet broadband, computers, and technological infrastructure to support the implementation of DAVE. Conditions are far from being the perfect ones, and planned PRR investments to promote schools' digital transition is still arriving to schools. Nevertheless, the window of opportunity that was opened with the *Next Generation EU* instrument could not be wasted and the Ministry of Education decided to take a step forward: To mitigate internet restrictions in schools, DAVE was designed to allow online and offline test administrations, with a *standalone* solution for schools with poor internet

connections. To mitigate the shortage of computers and technological infrastructure, PA were performed with two non-crossing shifts:

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

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

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

When it comes to stakeholders' perceptions about DAVE,

> (…) meetings held with head teachers, teachers and parents, everyone understood that large-scale digital testing will cause a drastic reduction in certain bureaucracies. Teachers understand this perfectly. In the marking process, for example, something that used to take two weeks to do, they can do it in half of the time, with many advantages in terms of reliability due to the process being online with the possibility of multiple teachers marking a given item (...). Still, we couldn't make everything go flawlessly and, naturally, there was also criticism.[3] (President of IAVE)

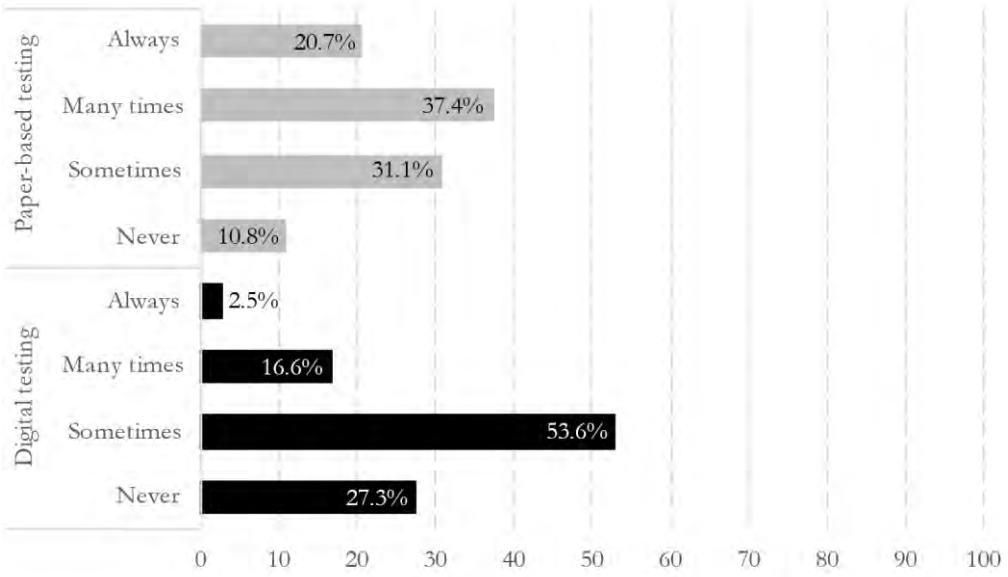
[3] See, e.g., ANPRI report (2023) [National Association of Informatic Teachers] available in Portuguese through https://www.anpri.pt/mod/forum/discuss.php?d=13742

RQ3. Teachers' ICT Use and Acceptability of DAVE

On the teachers' survey, it was possible to observe that only half of Portuguese teachers (52.7%) make a regular use of ICT in the classroom, such as the use of internet pages, PowerPoint©, videos, learning platforms and tools for the creation of digital educational resources. When it comes to classroom testing, as shown in Figure 2, 10.8% of teachers never use paper-based tests while 58.1% have a regular use of paper-based tests. As for digital testing, 27.3% of teachers never use digital based tests in classroom assessment and only 19.1% have a regular use of this kind of testing.

Figure 2

Paper-Based and Digital Testing in Classroom Assessment (%)



Source: Primary data, 2023.

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

Nevertheless, when considering the whole sample and which test(s) should be computer-based, teachers tend to reject the administration of the second, fifth and eighth grades PA on the computer (only 38.5%, 35.7% and 38.8%, respectively, agree with computer-based PA; as shown in Table 3). Hence, only 29.4% agree with 9th grade computer-based PFC, while 23.6% agree with 11th and 12th grades computer-based EN.

Among the teachers who agree with at least one computer-based test (i.e., within the subgroup of teachers who participated in the survey and agree with some test(s) being performed on a computer (n= 1 524; 57.0%)), they tend to support the existence of computer-based PA, with no stakes for students, as also shown in Table 3. In addition, within this subgroup of teachers that agree with some computer-based test(s), they still tend to support the existence of computer-based PFC in the ninth grade, which already involves some stakes for the conclusion of basic education. Although, these teachers tend to reject computer-based secondary education final exams (EN), which are used for the conclusion of secondary

education and higher education access. This means that within teachers who agree with some computer-based test(s), they tend to reject computer-based testing as stakes are getting higher.

Table 3

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

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

Note: On the sample (with $n = 2673$), a group of teachers (with $n = 1\,524$) agree with some kind of computer-based testing. On Table 3 it is reported the percentage when considering the whole sample; and the relative percentage when considering this subgroup of teachers only, i.e., within the group of teachers who agree with some kind of computer-based testing. Source: Primary data, 2023.

Discussion

The results showed that, on the one hand, some teachers and head teachers in Portugal consider that the schools are ready, and they agree with the implementation of DAVE. On the other hand, survey data showed that Portuguese teachers tend to reject the existence of computer-based large-scale tests. Further, many head teachers consider the digital transition is inevitable in the future, but schools are not ready yet for such a process due to a substantial lack of investment before its implementation.

First, concerns about schools' technological capacity to implement such a project arise as PRR investments have not yet arrived to schools. In the TALIS 2018 report, more than 60% of Portuguese lower secondary head teachers have already reported that the school's capacity to provide quality instruction is hindered "quite a bit" or "a lot" by the shortage or inadequacy of digital technology for instruction. Additionally, 35% also reported an insufficient internet access in schools (OECD, 2022). Although the COVID-19 pandemic has brought some technological improvements to schools, most of the interviewed head teachers reported that there is still an enormous lack of equipment and internet quality for such implementation as DAVE. Despite these restrictions, IAVE was able to successfully implement assessment tests (PA) in digital format in 2023, with different solutions being provided in tests design and administration to mitigate these limitations.

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

Portuguese head teachers about the gap widening between internal and external assessment practices.

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

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

Final Considerations

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

The process of evaluating acceptability of a testing initiative should consider specific circumstances that affect stakeholders' individual standards, and several other aspects, such as

applicable laws, regulations, and alternative measurement devices that are readily available (AERA et al., 2014). Portuguese teachers and head teachers have their own conceptions about external assessment purposes and practice (e.g., Cipriano & Martins, 2021; Machado et al., 2022), and this study showed that, for these stakeholders, there are additional powerful contextual circumstances affecting DAVE's trust and acceptability. Regarding applicable laws and regulations affecting acceptability, the debate about trust and acceptability on the use of computer-based technologies when stakes are involved is not new nor restricted to large-scale testing. For example, when it comes to trust and acceptability of e-voting, some countries have completely rejected the use of the internet on voting, while only a few countries have run a significant number of internet elections on a national scale. Hence, of these few countries, only Estonia has continued with plans for universal adoption (Gibson et al., 2016). While new laws and regulations can promote and support computer-based large-scale testing initiatives, decisions about the use and format of external assessment tests are always linked to the political and ideological choices of those who have the responsibility to put them into practice (Fernandes, 2019). In Portugal, it is a political decision.

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

Acknowledgements

The authors would like to express their gratitude to all the teachers, head teachers, and the president of IAVE for their participation in this research project.

References

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115-29. https://doi.org/10.1093/applin/14.2.115

Aloisi, C. (2023). The future of standardised assessment: Validity and trust in algorithms for assessment and scoring. *European Journal of Education*, *58*(1), 98-110. https://doi.org/10.1111/ejed.12542

Baker, E. L. (2001) Testing and assessment: A progress report. *Educational Assessment*, *7*(1), 1-12, https://doi.org/10.1207/S15326977EA0701_1

Bennett, R. E. (2001). How the Internet will help large-scale assessment reinvent itself. *Education Policy Analysis Archives*, *9*(5). https://doi.org/10.14507/epaa.v9n5.2001

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappa International*, *80*(2), 139–144. http://www.jstor.org/stable/20439383

Bryman, A. (2012). *Social research methods* (4th ed.). Oxford University Press.

Cipriano, G., & Martins, S. C. (2021). Beliefs on assessment and grade repetition among teachers in Portugal. *Meta: Avaliação*, *13*(39), 248–273. https://doi.org/10.22347/2175-2753v13i39.3033

Conselho Nacional de Educação (CNE). (2021). *Educação em tempo de pandemia: Problemas, respostas e desafios das escolas* [Education in pandemic times: problems, responses, and challenges of schools]. https://www.cnedu.pt/pt/publicacoes/estudos/1674-educacao-em-tempo-de-pandemia-problemas-respostas-e-desafios-das-escolas

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Sage Publications.

Direção-Geral da Administração Escolar [Directorate-General for School Administration]. (DGAE). (2019). *Códigos de agrupamentos de escolas / Escolas não agrupadas da rede do Ministério da Educação*. https://www.dgae.medu.pt/download/recrutamento/outros/201819/20190307_Rec_Codigos-AE_ENA-2019_2020.pdf

Direção-Geral de Estatísticas da Educação e Ciência [Directorate-General for Education and Science Statistics]. (DGEEC). (2023). *Perfil do docente 2021/2022*. https://www.dgeec.medu.pt/api/ficheiros/657206005f39ee77721e9b7f

Direção Regional de Administração Escolar da Madeira [Regional Directorate of School Administration of Madeira]. (DRAEM). (2023). *Quadros síntese estatísticas da educação*. https://www.madeira.gov.pt//Portals/16/Users/001/13/513/Quadros%20Síntese%20Estatísticas%20da%20Educação%202021_2022.pdf

Direção Regional de Educação e Administração Educativa dos Açores [Regional Directorate of Education and School Administration of Azores]. (DREAEA). (2023). *Estatísticas da educação*. https://edu.azores.gov.pt/wp-content/uploads/2023/09/ESTATISTICAS-DA-EDUCACAO-2021_22-1.pdf

Ercikan, K., Asil, M., & Grover, R. (2018). Digital divide: A critical context for digitally based assessments. *Education Policy Analysis Archives*, *26*(51). https://doi.org/10.14507/epaa.26.3817

European Commission (EU). (2020). *Portugal's recovery and resilience plan*. https://commission.europa.eu/business-economy-euro/economic-recovery/recovery-and-resilience-facility/portugals-recovery-and-resilience-plan_en

Eurydice. (2023). *National reforms in school education: Portugal.* European Commission. https://eurydice.eacea.ec.europa.eu/national-education-systems/portugal/national-reforms-school-education

Fernandes, D. (2013). Avaliação em educação: Uma discussão de algumas questões críticas e desafios a enfrentar nos próximos anos [Assessment in education: A discussion of some critical issues and challenges for the coming years]. *Ensaio*, *21*(78), 11–34. https://doi.org/10.1590/S0104-40362013005000004

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 A. Ferreira (Ed.), *Avaliação externa e qualidade das aprendizagens* (pp. 21-49). Conselho Nacional de Educação.

Fernandes, D. (2019). Avaliações externas e aprendizagens dos alunos: Uma reflexão crítica [External assessment and students learning: A critical reflection]. *Linhas Críticas.* https://doi.org/10.26512/lc.v25i0.24579

Geijsel, F., Sleegers, P., Van Den Berg, R., & Kelchtermans, G. (2001). Conditions fostering the implementation of large-scale innovation programs in schools: Teachers' perspectives. *Educational Administration Quarterly*, 37(1), 130–166. https://doi.org/10.1177/00131610121969262

Instituto de Gestão Financeira da Educação [Institute of Education Financial Management] (GesEdu). https://www.gesedu.pt/PesquisaRede

Gibson, J. P., Krimmer, R., Teague, V., & Pomares, J. (2016). A review of e-voting: The past, present and future. *Annals of Telecommunications*, *71*, 279-286. https://doi.org/10.1007/s12243-016-0525-8

Gipps, C., & Stobart, G. (2003). *Alternative assessment*. In T. Kellaghan & D. Stufflebeam (Eds), *International handbook of educational evaluation* (pp. 549-575). Kluwer. https://doi.org/10.1007/978-94-010-0309-4_33

Instituto de Avaliação Educativa –[National Institute of Educational Assessment] (IAVE). (2022). *O DAVE – Desmaterialização da avaliação externa*. https://iave.pt/wp-content/uploads/2022/04/1_DAVE-1.pdf

Johnson, M., & Shaw, S. (2018). What is computer-based testing washback, how can it be evaluated and how can this support practitioner research? *Journal of Further and Higher Education*. https://doi.org/10.1080/0309877X.2018.1471127

Kellagham, T., & Madaus, G. (2003). *External (public) examinations*. In T. Kellaghan & D. Stufflebeam (Eds), *International handbook of educational evaluation* (pp. 577-602). Kluwer. https://doi.org/10.1007/978-94-010-0309-4_34

Kerrey, B., & Isakson, J. (2000). *The power of the Internet for learning: Moving from promise to practice.* (Report). Web-Based Education Commission.

Linn, R. L. (2001). A century of standardized testing: Controversies and pendulum wwings. *Educational Assessment*, *7*(1), 29-38, https://doi.org/10.1207/S15326977EA0701_4

Machado, E. A., Flores, M. A., Pereira, D., Fernandes, E., & Costa, L. (2022). Políticas e práticas de avaliação externa dos alunos na perspetiva dos diretores: O caso das provas de aferição [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. https://doi.org/10.34632/investigacaoeducacional.2022.10958

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

Madaus, G., & Russell, M. (2010). Paradoxes of High-Stakes Testing. *Journal of Education,* 190, 21–30. https://doi.org/10.1177/0022057410190001-205

National Academy of Education (NAE). (2021). *Educational assessment in the COVID-19 era and beyond.* NAE. 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

Office of Qualifications and Examinations Regulation (Ofqual). (2020). *Online and on-screen assessment in high stakes, sessional qualifications: A review of the barriers to greater adoption and how*

these might be overcome. https://www.gov.uk/government/publications/online-and-on-screen-assessment-in-high-stakes-sessional-qualifications

Oldfield, A., Broadfoot, P., Sutherland, R., & Timmins, S. (2012). Assessment in a digital age: A research review. Graduate School of Education, University of Bristol. https://www.bristol.ac.uk/media-library/sites/education/documents/researchreview.pdf

Organisation for Economic Co-operation and Development (OECD). (2023). *Education at a glance 2023: OECD indicators*. OECD Publishing. https://doi.org/10.1787/e13bef63-en

Organisation for Economic Co-operation and Development (OECD). (2022). *Mending the education divide: Getting strong teachers to the schools that need them most*. TALIS, OECD Publishing. https://doi.org/10.1787/92b75874-en

Rädiker, S., & Kuckartz, U. (2020). *Focused analysis of qualitative interviews with MAXQDA: Step by step*. Maxqda Press. https://doi.org/10.36192/978-3-948768072

Recuperar Portugal. (2021) *Plano de recuperação e resiliência (PRR).* [Recovery and resilience plan]. https://recuperarportugal.gov.pt/2021/06/13/investimento-td-c20-i01/

Richardson, M., & Clesham, R. (2021). Rise of the machines? The evolving role of AI technologies in high-stakes assessment. *London Review of Education*, *19*(1), 0–13. https://doi.org/10.14324/LRE.19.1.09

Secretaria Geral da Educação e Ciência [General Secretariat of Education and Science] (SGEC). (2022). Investimento C20- i01 "Transição digital na educação." [Orientação téchnica]. https://recuperarportugal.gov.pt/wp-content/uploads/2022/02/Orientacao-Tecnica_SGEC_28_04_2022-Republicacao.pdf

Shaw, S., & Crisp, V. (2015). *Reflections on a framework for validation-Five years on.* Cambridge Assessment. https://www.cambridgeassessment.org.uk/Images/465780-reflections-on-a-framework-for-validation-five-years-on.pdf

Stobart, G., & Eggen, T. (2012). High-stakes testing - value, fairness and consequences. *Assessment in Education: Principles, Policy and Practice, 19*(1), 1–6. https://doi.org/10.1080/0969594X.2012.639191

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: grounded theory procedures and techniques.* Sage publications.

Tomas, C., Borg, M., & McNeil, J. (2015). E-assessment: Institutional development strategies and the assessment life cycle. *British Journal of Educational Technology*, *46*(3), 588-596. https://doi.org/10.1111/bjet.12153

Tsagari, D., & Cheng, L. (2017). Washback, impact, and consequences revisited. In: E. Shohamy, I. Or & S. May (Eds.), *Language testing and assessment. Encyclopedia of language and education*. Springer. https://doi.org/10.1007/978-3-319-02261-1_24

About the Authors

Gabriel Cipriano
Centro de Investigação e Estudos de Sociologia (CIES-Iscte), Iscte – Instituto Universitário de Lisboa, Portugal
gabriel.cipriano@iscte-iul.pt
https://orcid.org/0000-0002-0991-5499
Gabriel Cipriano holds a master's degree in School Administration, and he is a Ph.D. candidate in Policies of School Administration and Management at CIES-Iscte. Member of the Association for Educational Assessment - Europe, and the *Association pour le Développement des Méthodologies d'Évaluation en Éducation*. His main research interests are school administration, educational assessment, and school commuting policies.

Susana da Cruz Martins
Centro de Investigação e Estudos de Sociologia (CIES-Iscte), Iscte – Instituto Universitário de Lisboa, Portugal
susana.martins@iscte-iul.pt
https://orcid.org/0000-0002-5871-9849
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.

education policy analysis archives

Volume 32 Number 20 April 2, 2024 ISSN 1068-2341