



## OPEN ACCESS

## EDITED BY

Artem Artyukhov,  
Sumy State University, Ukraine

## REVIEWED BY

Khalida Parveen,  
Southwest University, China  
Sarfraz Aslam,  
UNITAR International University, Malaysia

## \*CORRESPONDENCE

Isabel Fialho  
✉ ifialho@uevora.pt

RECEIVED 12 January 2023

ACCEPTED 03 May 2023

PUBLISHED 26 May 2023

## CITATION

Fialho I, Cid M and Coppi M (2023) Grounding and improving assessment in higher education: a way of promoting quality education. *Front. Educ.* 8:1143356. doi: 10.3389/educ.2023.1143356

## COPYRIGHT

© 2023 Fialho, Cid and Coppi. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Grounding and improving assessment in higher education: a way of promoting quality education

Isabel Fialho\*, Marília Cid and Marcelo Coppi

Center for Research in Education and Psychology of the University of Évora, Évora, Portugal

Assessment continues to be one of the topics that most concern teachers and most dissatisfies students. In this paper we describe and evaluate the impact of the training course “Grounding and Improving Pedagogical Assessment in Higher Education,” based on a descriptive survey questionnaire, applied in a convenience-selected sample of 31 teachers from a public university in Portugal, who participated in the pedagogical training course. The questionnaire was applied before the beginning of the course (pre-test) and after its completion (post-test). Data were analyzed using statistical procedures that consisted of frequency analysis and non-parametric analysis, using the Mann–Whitney and Wilcoxon tests. The results revealed several changes in teachers’ conceptions of assessment and demonstrated the positive impact of the course, since the comparison of the pre and post-test results showed that relevant changes occurred towards the most current theories of pedagogical assessment, being more significant in the group of teachers without pedagogical training.

## KEYWORDS

pedagogical training, higher education, assessment, quality, education

## 1. Introduction

The quality of educational systems has been a priority on world political agendas, especially in recent decades. On January 1, 2016, the United Nations (UN) resolution entitled “Transforming our world: Agenda 2030 for Sustainable Development” took effect, consisting of 17 goals, broken down into 169 goals, which was approved by world leaders, on September 25, 2015, at a memorable summit at UN Headquarters in New York (USA). The nations involved agreed that the world could change and get better and accepted to improve people lives by the year 2030 by bringing together mainly governments, media, businesses, and institutions of higher education. One of the strengths of the Sustainable Development Goals (SDG) is their universal and multi-level perspective linking global, national, regional and local contexts in the sense of instilling a change that sees sustainable development as the prevailing paradigm to social transformation (Alcántara-Rubio et al., 2022).

Of the 17 SDGs that are part of the 2030 Education Agenda, under the responsibility of UNESCO, Goal 4 (Quality Education) highlights the importance of education for sustainable development, being a core goal that cuts across all other goals, underlying the importance of “Ensuring inclusive and equitable quality education, and promoting lifelong learning opportunities for all.” In this context, higher education institutions assume great responsibility as teaching, learning and research organizations, which ensure the initial and continuous

training of citizens and play a fundamental role in the generation and dissemination of knowledge.

The implementation of the SDGs in higher education has come a long way since then and a few studies have come to realize this. According to [Alcántara-Rubio et al. \(2022\)](#), the most recent literature reveals that the integration of the SDGs in the curriculum of higher education institutions is being done, but its implementation is still scarce, especially in terms of the ability to integrate the SDGs in a participatory, community and practical way. Among the SDGs, SDG4 was identified as the objective most present, which is justified by the fact that education is central and is strongly related to all other objectives as it plays a fundamental role in the achievement of the Agenda as a whole. As a matter of fact, in a study carried out in 2019, which aimed to analyze the integration of the SDGs in Portuguese public higher education institutions, namely at the level of undergraduate and masters courses, it was found that less than 10% of the 2,556 undergraduate and master's courses analyzed directly addressed at least one SDG ([Aleixo et al., 2019](#)). More recently, [Chaleta et al. \(2021\)](#) carried out a mapping study of the Sustainable Development Goals in the curricular units of the degree courses at the School of Social Sciences at the University of Évora (Portugal), having verified that all curricular units with students enrolled in 2020/2021 ( $n = 374$ ) made reference to the Sustainable Development Goals and that SDG4 (Quality Education) was common to all curricular units. These results are in line with others such as [Salvia et al. \(2019\)](#) who concluded that SDG 4 was one of the most highlighted objectives by participating experts from different geographic regions, in a total of 266 from North America, Latin America/Caribbean, Africa, Asia, Europe and Oceania.

Although these results are somewhat positive, the conclusions of other studies show that there is still much to be done in order to integrate education for sustainability in higher education institutions (HEIs), since its transformative role is yet to be achieved ([Alcántara-Rubio et al., 2022](#)). It should be noted that education for sustainable development includes quality education and implies the acquisition of skills that allow responding to the challenges facing humanity today. Thus, curriculum plans cannot fail to express concerns about pedagogical innovation, incorporating student-centered pedagogical models, oriented towards action, participation and collaboration in problem solving, with the development of projects that require critical thinking and reflection ([UNESCO, 2017](#)).

In terms of quality education, the European Higher Education Area, following the Bologna process, strengthened the need for quality education and drove the creation of national and supranational systems and agencies for the evaluation and accreditation of training and processes for promoting and guaranteeing quality. The European Standards and Guidelines (ESG) are references for the assessment and accreditation guidelines followed by agencies and HEIs in each member state and are divided into three parts: (1) Internal quality assurance, (2) External quality assurance, and (3) Quality assurance agencies. The ESG for internal quality assurance, in particular standards 1.5 Teaching staff, point out that HEIs should provide and promote initiatives and opportunities for the professional development of teaching staff and encourage innovation in teaching methods and the use of new technologies. Standard 1.5 Teaching staff stands that "Institutions should assure themselves of the competence of their teachers" and reinforces that "the teacher's role is essential in creating a high quality student experience and enabling the acquisition of

knowledge, competences and skills (...). Higher education institutions have primary responsibility for the quality of their staff and for providing them with a supportive environment that allows them to carry out their work effectively" ([ESG, 2015](#), p. 11). Within this framework, four guidelines aim to guarantee the quality of teaching performance:

- sets up and follows clear, transparent and fair processes for staff recruitment and conditions of employment that recognize the importance of teaching;
- offers opportunities for and promotes the professional development of teaching staff;
- encourages scholarly activity to strengthen the link between education and research;
- encourages innovation in teaching methods and the use of new technologies ([ESG, 2015](#), p. 11).

It appears that the culture of quality, with regard to the pedagogical dimension, is a priority in the political agendas of European HEIs, which have instituted various mechanisms to promote the quality of the teaching, learning and assessment processes. Some initiatives stand out, namely in Sweden where pedagogical training measures were adopted; in the Netherlands, where the University Teaching Qualification (UTQ) was introduced with the aim of certifying the pedagogical skills of teachers, combining the frequency of pedagogical training with professional performance; in Denmark, where all HEIs have teaching centers and organize training programs; in the United Kingdom, where the Higher Education Academy (HEA) developed the UK Professional Standards Framework (UKPSF), which consists of a teaching-learning support guide that brings together a set of guidelines for good performance in the exercise of teaching activity. Following the European trend, courses in Portugal are now subject to periodic external evaluation (by A3ES) based on various indicators, namely "the qualification and adequacy of the teaching staff [and] the strategy adopted to guarantee the quality of teaching- learning-assessment" ([Pereira and Leite, 2020](#), p. 138). The Bologna process generated a strong increase in the quality of teaching, in line with the European agenda for improving the quality of teachers ([European Commission, 2010](#)). These two interconnected aspects have drawn attention of higher education institutions, constituting a priority and a strategic imperative for the competitive success of universities ([Fialho and Cid, 2021](#)).

Considering the competences attributed to HEIs with regard to university autonomy, they have the responsibility of promoting continuous improvement in the quality of the teaching staff. With the aim of guaranteeing the permanent scientific updating of teachers, both statutes for teaching careers (University Higher Education and Polytechnic Higher Education) encourage the strengthening of the link between teaching and research. However, the tendency has been to make teaching secondary in favor of research [[Federação Académica do Porto \(FAP\), 2021](#)]. Being a teacher in higher education seems to have less importance than being a researcher in such a way that training in the specific scientific area is, in general, more valued than pedagogical training. Therefore, "the pedagogical preparation of higher education teachers is now a social and professional imperative" ([Almeida, 2020](#), p. 4).

Bologna's teaching paradigm points to more participatory processes, which increase the student's sense of commitment,

autonomy and responsibility in the learning process and which place the teacher in the role of facilitator and mediator. It is a shift from a teaching-centered paradigm to a student-centered paradigm, going against the instructional teaching culture strongly rooted in universities that explains the fact that many teachers do not consider pedagogical training a priority in their professional development (Fialho and Cid, 2021).

Students in the 21st century have nevertheless demands that the teaching staff of higher education institutions can hardly meet without extensive and plural training. Today's students are immersed in the digital world, with very diverse cultural and social origins, very different profiles and paths. Therefore, the expression "the teacher is the key to success" takes on new meanings, having solid scientific knowledge and high investigative skills is not enough to meet the demands of students in this century, the path to success is made with pedagogical innovation, with approaches in contexts that mobilize significant learning with impact and quality, providing constant feedback, collaboration and reflection. It is about replacing the expository model, centered on the teacher, based on dependence on encyclopaedic knowledge, for a model centered on the student, requiring autonomy, responsibility and new skills: the ability to research and select, interpret and analyze information, differentiate points of view, to have a critical sense and creativity.

Higher education institutions are thus subject to increasing pressure for pedagogical innovation, whether from new students or from the social and economic world. The traditional missions of Higher Education systems – teaching, researching and providing services to the community – remain valid but, in this context of globalization, uncertainties and challenges, is reinforced its mission of training citizens capable of integrating and participating in increasingly complex and demanding social and professional environments, developing skills for lifelong learning, which implies a reconfiguration of teaching. "The content (what is taught), the pedagogy (how it is taught) and the way technology is integrated into the classroom (framework) become part of the domains to which the teacher has to respond" [Federação Académica do Porto (FAP), 2021, p. 49].

Improving pedagogical practice "goes through investing, intentionally, systematically and coherently in training processes that involve teachers in the permanent questioning of practice and in the dialectic between theory and practice" (Almeida, 2020, p. 20), appealing to mechanisms that ensure continuous updating of knowledge and pedagogical skills, many of which in the digital area. But it is also essential to invest in support services for the teaching activity, namely in supporting the use of educational technologies, tools and digital platforms.

More and more HEIs have pedagogical training programs for their teachers. In 2014, the Foundation for Science and Technology opened a public tender to finance projects aimed at sharing and disseminating pedagogical innovation experiences in Portuguese higher education, with 181 applications being submitted (Ministério da Educação e Ciência, 2015). However, along with important changes in higher education pedagogy, practices still do not meet the pedagogical trends advocated by Bologna, as evidenced by some studies carried out with higher education teachers. The study conducted by Pereira and Leite (2020) underlines that the "changes made, although relevant, are not yet fully in line with the desired levels

of quality within the framework of the quality agenda associated with the BP [Bologna Process]" (p.148).

On the other hand, the study of Flores et al. (2007) reveal that the vast majority of teachers (76.5%) attended training related to teaching to the detriment of those related to research and more than 85% refer as motivations for attending training, the development of skills, professional development, deepening practical issues related to pedagogical activity, updating knowledge and reflecting on practice.

The demands and challenges are unquestionable, however, higher education institutions have taken on pedagogical training in different ways, although many are on a path of consolidation, others seem not to give it much attention yet (Xavier and Leite, 2019). Therefore, the guarantee of pedagogical quality will have to go through "the legitimacy conferred on teacher training, through the commitment collectively assumed by the institution" (Ogawa and Vosgerau, 2019, p. 12). Aware of this, many HEIs have planned and organized pedagogical training for their teaching staff, with training in assessment being one of the most requested, regarded as a recurrent training need. In fact, research in pedagogical assessment has accumulated studies since the 1990s, experiences and evidence that point to the need of prioritizing assessment strategies that help students learn rather than simply classifying their learning, and that formative assessment is the one that has the most positive effects on learning (Black et al., 2011; Black, 2013; Wiliam, 2017; Fialho et al., 2020). However, pedagogical assessment in higher education follows in general summative approaches centered on the use of tests and exams as a privileged means of assessing and classifying students.

Pedagogical assessment involves the collection, analysis, discussion and use of information regarding students' learning and its main purpose is to help students learning. It is important, therefore, to use different processes for collecting information insofar as students develop the various learning and skills included in the curriculum. This diversification of processes must take into account, not only the subjective nature of the assessment, but also the different students styles and ways of learning, so that formative assessment must prevail in assessment practices in order to provide high quality feedback to students (Brookhart, 2008; Fernandes, 2021a).

Formative assessment takes place on a daily basis in classes and is integrated into the teaching and learning processes, thus being associated with forms of regulation and self-regulation of those two processes. In order to guarantee that formative assessment has a real impact on learning, the proper use of feedback is required, since it informs students of what they are expected to learn, where they are in the learning process and what they need to do to learn (Wiliam, 2011; Fernandes, 2019a).

Summative assessment, on the other hand, does not follow the daily teaching practice in such a systematic way and allows for a balance or status of what the students have learned at the end of a certain time, it is therefore punctual and does not have the continuous nature of formative assessment. Thus, one of the purposes of summative assessment is to collect information that allows judging what students have learned, with or without classification (Popham, 2017). When we are dealing with a classificatory summative assessment, the information collected allows us to report the grades that are used to communicate what the students were able to achieve in relation to what is expected (Fernandes, 2019b).

The European agencies for promoting and validating the quality of HEIs have encouraged the questioning of the way(s) in which

teaching, learning and assessment are carried out. Returning to standards 1.3 Student-centered learning, teaching and assessment (ESG, 2015), it establishes that the “Institutions should ensure that the programs are delivered in a way that encourages students to take an active role in creating the learning process, and that the assessment of students reflects this approach” (ESG, 2015, p. 9) and presents seven guidelines that emphasize the student-centered teaching and learning process and highlight the role of assessment in the process self-regulation:

- respects and attends to the diversity of students and their needs, enabling flexible learning paths;
- considers and uses different modes of delivery, where appropriate;
- flexibly uses a variety of pedagogical methods;
- regularly evaluates and adjusts the modes of delivery and pedagogical methods;
- encourages a sense of autonomy in the learner, while ensuring adequate guidance and support from the teacher;
- promotes mutual respect within the learner-teacher relationship;
- has appropriate procedures for dealing with students’ complaints (ESG, 2015, p. 9).

The importance of assessment is highlighted in its formative aspect through the definition of specific guidelines for pedagogical assessment:

- Assessors are familiar with existing testing and examination methods and receive support in developing their own skills in this field;
- The criteria for and method of assessment as well as criteria for marking are published in advance;
- The assessment allows students to demonstrate the extent to which the intended learning outcomes have been achieved. Students are given feedback, which, if necessary, is linked to advice on the learning process;
- Where possible, assessment is carried out by more than one examiner;
- The regulations for assessment take into account mitigating circumstances;
- Assessment is consistent, fairly applied to all students and carried out in accordance with the stated procedures;
- A formal procedure for student appeals is in place (ESG, 2015, pp. 9–10).

In the study carried out by the Federação Académica of Porto, in the post-pandemic, the students point out that assessment was the worst aspect in the distance learning process experienced in the two academic years of the pandemic. They mentioned that the assessment methods mobilized were, for the most part, the same, or at least very similar to those used during face-to-face teaching. The opinion of teachers differs from that of students, as more than 70% of professors declared that they had modified the assessment methods and 41% stated that they had used innovative methods during the online teaching period [Federação Académica do Porto (FAP), 2021]. However, pedagogical innovation is not limited to the introduction of digital tools in the teaching-learning process, the transition to a teaching model that mobilizes the use of new technologies is inevitable given the “greater propensity of younger generations to use platforms and digital tools” [Federação Académica do Porto (FAP), 2021, p. 47].

From a list of 14 assessment tools, the most frequent response from students was final exams, even at a distance, among the 3,195 students surveyed, 81% said they had completed at least one final exam per discipline, of which 42% evaluate the experience negatively. Dissatisfaction with the assessment processes and the perception of unfairness regarding the relationship between subject knowledge and assessment results can perpetuate anxiety and stress that impact school success and reinforce the need to rethink assessment practices [Federação Académica do Porto (FAP), 2021].

The pandemic had unquestionable effects on the pedagogical process. It will have contributed to an acceleration of the education systems digitalization process that was already in progress, to the reinforcement of the students’ autonomy, mobilizing and involving them in a more active way in the learning process, for new dynamics of work, for the quality of feedback and for the diversity of assessment methods and the enhancement of continuous assessment regimes, in which assessment is an integral part of the learning process [Federação Académica do Porto (FAP), 2021].

Despite these changes and possible positive effects that resulted from the conditions that educational institutions have been experiencing in recent times, it is often verified that assessment continues to be one of the topics that most concern teachers and that raises more dissatisfaction in students (Flores et al., 2021). Thus, the preparation of teachers in this area remains on the agenda, with the training reported in this article, which had as its objectives (1) to identify higher education teachers’ conceptions of assessment and (2) evaluate the effect of the training in educational assessment on their conceptions, being an example among those that have been carried out in many different institutions of higher education. Based on the arguments previously presented, it was considered that training should reinforce and deepen the grounding necessary for changing and improving assessment practices.

## 2. Method

The information presented here, is part of an exploratory study, essentially quantitative in nature, through a descriptive survey, stems from data obtained from a questionnaire applied to teachers at a public university in Portugal who participated in a pedagogical training course focused on assessment. The questionnaire was applied in two moments: before the beginning of the course (pre-test) and after its completion (post-test), with the purpose of collecting information about conceptions and assessment practices of higher education teachers.

### 2.1. The training

This training started in 2021 in the form of a course named “Grounding and Improving Pedagogical Assessment in Higher Education,” which adopted a training model based on the idea of a process that develops in a reflective spiral of successive cycles of planning, action, and evaluation of the outcome of the action (Kemmis and McTaggart, 1992). It was intended for university teachers and aimed to conceptualize and ground the practices in order to improve pedagogical assessment.

Three editions were held, the first two lasting 14 h and the following 16 h. The course took place entirely in a digital environment

with synchronous (on the Zoom platform) and asynchronous (on the Moodle platform) sessions lasting 1 h each. Throughout the sessions were proposed different types of tasks in groups or individually (in synchronous and asynchronous sessions), all with feedback, either immediate feedback (self-corrective), written feedback or oral feedback provided by the researchers/trainers in synchronous sessions.

Teachers who participated in the trainings answered a questionnaire, available online on the LimeSurvey platform, at the beginning and at the end of the course. By answering the questionnaire, the teachers gave their informed consent, agreeing to participate in the study, understanding that their participation was voluntary and being aware that the data would be collected anonymously, according to the rules of the data protection commission.

## 2.2. The instrument

The questionnaire was developed by the authors as part of the training course in question and consists of three parts/groups of questions. Part I includes questions for the respondents' socio-professional characterization, Part II includes questions aimed at learning about assessment practices, and Part III includes questions about conceptions within the scope of pedagogical assessment. In this article we present the data from the socioprofessional characterization and the data that were collected in question 9 of Part III.

Question 9 consists of 37 statements (items), with response options on a 4-point Likert-type agreement scale (1 = Strongly Disagree; 2 = Disagree; 3 = Agree; and 4 = Strongly Agree). These items were distributed into four categories: (1) Conceptions of pedagogical assessment (10 items), which encompasses issues related to the purpose, credibility, criteria, and subjectivity of assessment, and the quality and diversity of assessment instruments; (2) Conceptions of summative assessment and formative assessment (19 items), whose topic refers to the potentiality, goals, and usefulness of these two assessment modalities; (3) Conceptions of feedback (5 items), which addresses the quality and usability of feedback types; (4) Conceptions of learning grading (3 items), which discusses the composition of the final learning grading.

## 2.3. Participants' characterization

The sample consists of a convenience-selected sample (Ghigione and Matalon, 1992; Hill and Hill, 2005) of 31 university teachers who participated in at least one of the pedagogical training course. It is important to make it clear that is an exploratory study and due to its size, the sample is not representative, and therefore the results are not generalizable to the population of professors at the university where the research was carried out.

The majority of the teachers of the sample (45.2%) were between 50 and 59 years old, with an average of 20.13 years of service (SD = 10.80), of which 80.6% are female and 19.4% are male. Regarding the distribution to schools of the university,<sup>1</sup> 45.2% of the

teachers belong to the School of Science and Technology, 41.9% to the School of Social Sciences, and 12.9% to the School of Arts. Of the 31 teachers who signed up for the course, 32.3% had pedagogical training, 36.4% were trained in Initial Training, 45.5% in Continuing Training, and 18.2% in both.

It is important to stress that 48.4% of the teachers said they felt constrained/difficulty in the scope of pedagogical assessment and that only 10 teachers had undergone some form of pedagogical training previously.

## 2.4. Data analysis procedure

Data were analyzed using statistical procedures that consisted of frequency analysis and, due to the nature of the sample, non-parametric analysis, using the Mann-Whitney and Wilcoxon tests, using the SPSS v. 27 software. The Mann-Whitney test was used to compare the mean scores of agreement between the group of teachers who had pedagogical training before the course and those who did not have this training, regarding the four categories of conceptions about pedagogical assessment (Field, 2009). The Wilcoxon test was applied to: (a) compare the mean agreement of teachers in the four categories of conceptions of pedagogical assessment in the pre and post-test; (b) compare the mean agreement of teachers with previous pedagogical training in the pre and post-test; and (c) compare the mean agreement of teachers with no previous pedagogical training in the pre and post-test (Field, 2009).

## 3. Results and discussion

In order to meet the objectives of the study, we present the answers to the questions included in Part III of the questionnaire. In order to present the results in a logical and sequential manner, we will describe the analyses by category.

### 3.1. Conceptions of pedagogical assessment

Regarding the conceptions about pedagogical assessment, it is observed that there were statistical differences between the pre and post-test in statements 1 and 5 (Table 1; Figure 1). In both statements, the degree of agreement in the pre-test was higher than in the post-test, a result that, due to the course objectives and components, was expected. Regarding statement 1 (The main purpose of pedagogical evaluation is to grade learning), the mean agreement of 2.52 (SD = 0.72) decreased to 1.58 (SD = 0.56) in the post-test, demonstrating that teachers understood that the main purpose of instructional evaluation is broader than just grading learning.

Before the course, for the same statement, the mean agreement of teachers who did not have pedagogical training was statistically higher ( $U = 51.000$ ;  $p < 0.05$ ) than that of those who did. The mean for the former group of teachers was 2.00 (SD = 0.81), while the mean for the

<sup>1</sup> The University where the study was conducted was organized into four major organic units: School of Social Sciences, School of Sciences and

Technology, School of Arts, and School of Nursing.

TABLE 1 Differences between the means of all teachers' conceptions of pedagogical assessment in the pre and post-test ( $n = 31$ ).

Statement	Test	Mean	Std. deviation	Z	p
1. The main purpose of pedagogical evaluation is to grade learning.	Pre-test	2.52	0.72	-4.010	0.000***
	Post-test	1.58	0.56		
2. The main purpose of pedagogical assessment is student learning.	Pre-test	3.35	0.71	-1.213	0.225
	Post-test	3.52	0.51		
3. The diversity of tasks and assessment instruments contributes to a fairer assessment.	Pre-test	3.74	0.44	-1.155	0.248
	Post-test	3.58	0.67		
4. The credibility of the assessment depends on the instruments.	Pre-test	3.06	0.85	-1.334	0.182
	Post-test	2.84	0.64		
5. The quality of the instruments guarantees the objectivity of the assessment.	Pre-test	3.00	0.93	-2.923	0.003**
	Post-test	2.39	0.88		
6. Teachers should discuss the assessment criteria with students.	Pre-test	3.39	0.76	-0.802	0.423
	Post-test	3.52	0.57		
7. Assessment criteria help students improve their learning.	Pre-test	3.39	0.67	-1.414	0.157
	Post-test	3.58	0.5		
8. The assessment criteria make the evaluation transparent.	Pre-test	3.68	0.47	-1.807	0.071
	Post-test	3.45	0.57		
9. Students must participate in the assessment process.	Pre-test	3.26	0.68	-0.566	0.572
	Post-test	3.35	0.66		
10. The assessment is always subjective.	Pre-test	2.71	0.78	-1.310	0.190
	Post-test	2.97	0.95		

\*\* $p < 0.01$ .\*\*\* $p < 0.001$ .

latter group was 2.76 (SD = 0.54). After the course, the mean agreement scores of both groups decreased (M = 1.40; SD = 0.52 and M = 1.67; SD = 0.57, respectively). However, only the decrease in the mean for the group of teachers who had no previous pedagogical training was significant ( $Z = -3.51$ ;  $p < 0.001$ ).

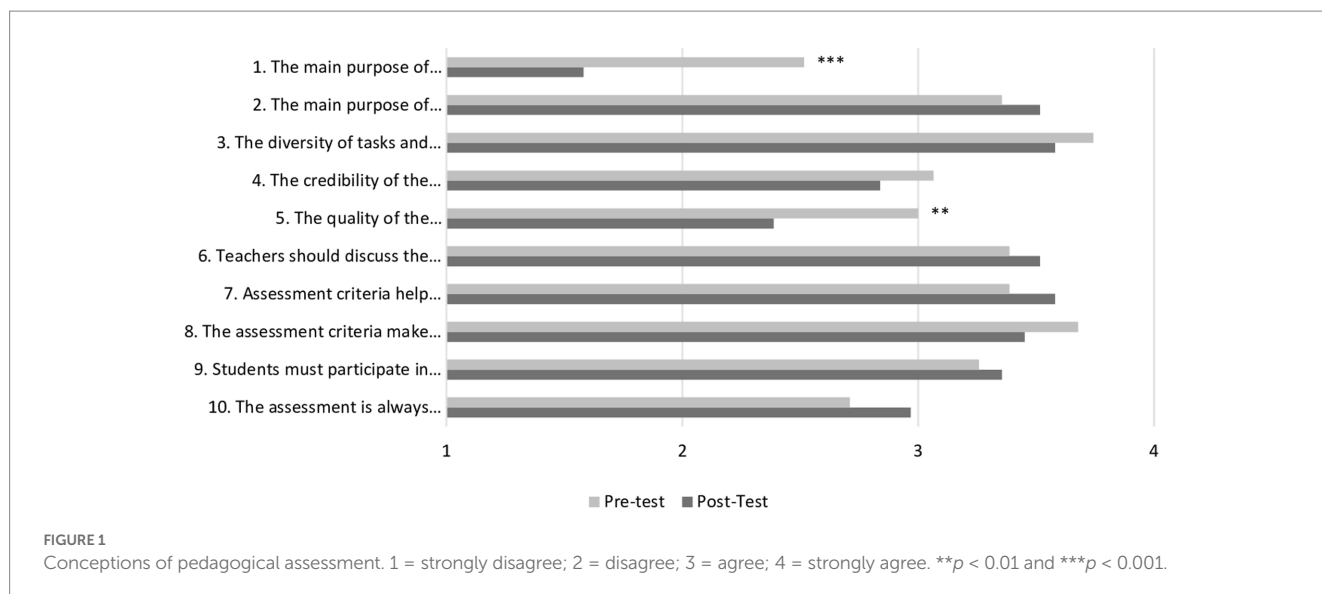
The change of conception observed presents itself as fundamental for the evaluation of the course impact, since the content of the statement is one of the key issues in the area of assessment. Since the 1990s, research has accumulated studies, experiences and evidence that point to the need to privilege assessment strategies that help students learn instead of simply grading their learning (Fialho et al., 2020). Indeed, the main purpose of pedagogical assessment is the gathering of information that allows the teacher to make good educational decisions in order to improve student learning (Popham, 2017). This gathering is related to several purposes that include "establishing classroom balance, planning and conducting instruction, placing students, providing feedback and incentives, diagnosing student problems and disabilities, and judging and grading academic learning and progress" (Russell and Airasian, 2008, p. 5).

It is worth noting that, before the course, there was also a significant difference ( $U = 53.000$ ;  $p < 0.05$ ) between the mean agreement of the two groups for statement 7 (Assessment criteria help students improve their learning). Both showed positive agreement means, that is, between the points of agree and strongly agree, with the mean of teachers who had higher pedagogical training (M = 3.80; SD = 0.42) to that of those who did not (M = 3.19; SD = 0.68). After the course, the means remained on the same spectrum of agreement

(M = 3.70; SD = 0.48 and M = 3.52; SD = 0.51, respectively) with no statistical difference between the groups.

It can be seen that teachers' conceptions of the importance of assessment criteria in improving student learning are in line with the literature, which states that assessment criteria assist in improving learning by allowing students "to focus their attention on the characteristics that will be assessed and to understand the quality of performance that is expected" (Russell and Airasian, 2008, p. 229). The assessment criteria should be made available and clearly expressed, constituting a fundamental element of students' orientation, ensuring the transparency of the assessment process (Fernandes, 2004). Teachers can share the criteria, "for example, by giving students a copy of the scoring rubrics you will use to evaluate their final work" or show "some examples over a range of quality levels and let the students figure out what is 'good' about the good work" (Nitko and Brookhart, 2014, p. 95).

About statement 5 (The quality of the instruments guarantees the objectivity of the assessment), the average of agreement of all teachers is 3.00 (SD = 0.93) in the pre-test, decreasing to 2.39 (SD = 0.88) in the post-test. In this case, it is observed that the post-test mean, although lower than the first, it still remained between the levels of disagree and agree, indicating the existence of belief in the objectivity of the assessment and that it depends on the quality of the instruments. Although there were no significant differences between the mean levels of agreement of the group of teachers who had pedagogical training and those who did not, both in the pre and post-test, there was a decrease in the mean level of agreement of those who already



had training from 2.50 (SD = 1.27) to 2.20 (SD = 0.36), and of those who did not have training from 3.24 (SD = 0.62) to 2.48 (SD = 0.98).

Despite the increase in disagreement, the non-existence of significant differences between teachers with and without pedagogical training points to the need for further clarification and reflection on the issue of objectivity. Assessment, as it is not an exact or objective science or a mere measurement, is a process imbued with subjectivity (Fernandes, 2021b). From this perspective, no instrument is able to provide information and produce exact, objective or bias-free measures of students' learning and skills. For this reason, the quality of the instruments, a fundamental characteristic, associated with the definition and sharing of assessment criteria, is able to reduce the subjectivity of the assessment process, but does not guarantee its complete objectivity (Fernandes, 2021b).

It is also noteworthy that when comparing the intragroup means in the pre and post-test, significant differences were observed in statements 1 (The main purpose of pedagogical evaluation is to grade learning), 4 (The credibility of the assessment depends on the instruments), and 5 (The quality of the instruments guarantees the objectivity of the assessment), only among teachers without pedagogical training (Table 2). For all three statements, the mean agreement of teachers in the post-test was statistically lower than in the pre-test.

The difference observed between the mean scores of agreement for statements 1 and 5 reinforce the discussion held earlier. In the case of statement 4 (The credibility of the assessment depends on the instruments), it is observed that while the mean of teachers with pedagogical training had a slight increase, from 2.90 (SD = 1.20) to 3.00 (SD = 0.67) establishing themselves at the agree point of the scale, for teachers without pedagogical training, the mean decreased from 3.14 (SD = 0.65) to 2.76 (SD = 0.63), with no significant differences between the means of the groups in the post-test. These results seem to reinforce the idea that teachers give great importance to instruments, although credibility is not only associated with the instrument itself, but also with the diversity and transparency of the whole evaluation process, in which "the criteria, purposes, procedures, moments, actors and processes of information gathering to be used should be known by the main actors" (Fernandes, 2021b, p.15).

### 3.2. Conceptions of summative assessment and formative assessment

Regarding conceptions of summative and formative assessment, it can be observed that there were significant differences between the pre and post-test in more than half of the statements, 11 out of 19 (Table 3; Figure 2). It can be seen that after the course, the teachers' mean agreement was statistically higher for statements 17, 22, 23, 24, 25, and 29 and significantly lower for statements 13, 14, 15, 26, and 27.

Among the statements that obtained significantly higher mean agreement in the post-test, it was found that, although all of them were already in the positive spectrum of the scale, that is, between the points of agree and strongly agree, in the pre-test, the mean agreement increased, bringing them closer to the point of strongly agree. In this sense, taking into account that all statements are in accordance with the currently accepted good assessment practices, it can be inferred that the course has positively interfered in the conceptions about summative and formative assessment.

It is worth mentioning the statements whose average agreement of all teachers in the post-test was statistically lower than the pre-test – statements 13 [Assessment tests (exams and tests) allow to measure learning objectively] allow for objective measurement of learning), 14 (Grading helps students become aware of their difficulties and the means to overcome them), and 15 (The grading of learning is useful for regulating teaching practices) – whose mean was between the disagree and agree spectrum. These data corroborate the study by Barreira et al. (2017), in which teachers agree that assessment is used to guide students to learn better, but disagree that the information resulting from assessments is used by students to guide and/or reorient their ways of studying.

In the case of statement 13, this decrease may point to a tendency to reduce the "culture of testing" (Pereira et al., 2015), with the eventual devaluation of this instrument in relation to others and the belief that assessment can be objective, as confirmed by studies in which teachers consider tests and/or final exams to be the best way to find out what students know and are capable of doing (Barreira et al., 2017). As discussed earlier, all assessment is subjective by nature, given that "assessment is inexorably associated with the perspectives,

TABLE 2 Differences between the means of the conceptions of teachers with and without pedagogical training about pedagogical assessment in the pre and post-test ( $n = 31$ ).

Statement	Pedagogical training	Test	Mean	Std. deviation	Z	p
1	Yes	Pre-test	2.00	0.82	-1.897	0.058
		Post-test	1.40	0.52		
	No	Pre-test	2.76	0.54	-3.508	0.000***
		Post-test	1.67	0.58		
2	Yes	Pre-test	3.60	0.70	-2.724	0.564
		Post-test	3.50	0.53		
	No	Pre-test	3.24	0.70	-1.604	0.109
		Post-test	3.52	0.51		
3	Yes	Pre-test	3.90	0.32	-2.000	0.414
		Post-test	3.60	0.97		
	No	Pre-test	3.67	0.48	-0.816	0.414
		Post-test	3.57	0.51		
4	Yes	Pre-test	2.90	1.20	-0.264	0.792
		Post-test	3.00	0.67		
	No	Pre-test	3.14	0.65	-2.138	0.033*
		Post-test	2.76	0.63		
5	Yes	Pre-test	2.50	1.27	-0.965	0.335
		Post-test	2.20	0.63		
	No	Pre-test	3.24	0.62	-3.025	0.002**
		Post-test	2.48	0.98		
6	Yes	Pre-test	3.50	0.85	-0.707	0.480
		Post-test	3.30	0.68		
	No	Pre-test	3.33	0.73	-1.355	0.175
		Post-test	3.62	0.50		
7	Yes	Pre-test	3.80	0.42	-0.577	0.564
		Post-test	3.70	0.48		
	No	Pre-test	3.19	0.68	-1.807	0.071
		Post-test	3.52	0.51		
8	Yes	Pre-test	3.70	0.48	-0.816	0.414
		Post-test	3.50	0.53		
	No	Pre-test	3.67	0.48	-1.667	0.096
		Post-test	3.43	0.60		
9	Yes	Pre-test	3.30	0.82	-0.378	0.705
		Post-test	3.20	0.79		
	No	Pre-test	3.24	0.62	-0.973	0.331
		Post-test	3.43	0.60		
10	Yes	Pre-test	2.80	0.92	-0.333	0.739
		Post-test	2.90	1.10		
	No	Pre-test	2.67	0.73	-1.355	0.176
		Post-test	3.00	0.89		

\* $p < 0.05$ .

\*\*\* $p < 0.001$ .

**TABLE 3** Differences between the means of all teachers' conceptions of summative assessment and formative assessment in the pre and post-test ( $n = 31$ ).

Statement	Test	Mean	Std. deviation	Z	p
11. The purpose of summative assessment is to grade the learning.	Pre-test	3.00	0.51	-0.799	0.425
	Post-test	2.87	0.85		
12. Assessment tests (exams and tests) measure what students know and are able to do at a given moment.	Pre-test	2.9	0.70	-1.334	0.182
	Post-test	3.13	0.81		
13. Assessment tests (exams and tests) allow to measure learning objectively.	Pre-test	2.45	0.62	-2.457	0.014*
	Post-test	2.03	0.66		
14. Grading helps students become aware of their difficulties and the means to overcome them.	Pre-test	2.9	0.65	-2.310	0.021*
	Post-test	2.48	0.81		
15. The grading of learning is useful for regulating teaching practices.	Pre-test	3.06	0.44	-2.668	0.008**
	Post-test	2.71	0.78		
16. Learning grading is useful for regulating learning.	Pre-test	2.94	0.57	-1.259	0.208
	Post-test	2.71	0.78		
17. Summative assessment can be used for formative purposes.	Pre-test	30.16	0.34	-2.066	0.039*
	Post-test	3.45	0.62		
18. Formative assessment should be used systematically, in the course of lessons.	Pre-test	3.26	0.63	-1.882	0.060
	Post-test	3.55	0.57		
19. Formative assessment is any assessment that takes place in the classroom.	Pre-test	2.84	0.78	-1.938	0.053
	Post-test	3.42	0.99		
20. Formative assessment should help students become aware of their difficulties and the means to overcome them.	Pre-test	3.61	0.49	-1.667	0.096
	Post-test	3.77	0.42		
21. The information gathered in the formative assessment is useful for the final grading of learning.	Pre-test	3.00	0.86	0.000	1.000
	Post-test	3.00	0.86		
22. The information gathered in formative assessment is useful for regulating teaching practices.	Pre-test	3.45	0.57	-2.840	0.005**
	Post-test	3.81	0.40		
23. The information gathered in formative assessment is useful for regulating learning.	Pre-test	3.45	0.51	-2.496	0.013*
	Post-test	3.74	0.51		
24. Self-evaluation promotes the improvement of learning.	Pre-test	3.13	0.76	-2.066	0.039*
	Post-test	3.42	0.62		
25. Hetero-assessment promotes the improvement of learning.	Pre-test	3.03	0.75	-2.829	0.005**
	Post-test	3.45	0.51		
26. Summative assessment is more rigorous than formative assessment.	Pre-test	2.29	0.64	-3.231	0.001**
	Post-test	1.61	0.61		
27. Summative and formative assessment are distinguished by the instruments used to gather information.	Pre-test	2.52	0.68	-2.920	0.004**
	Post-test	1.97	1.01		
28. Formative assessment is distinguished from summative assessment by the absence of grading.	Pre-test	2.19	0.70	-0.588	0.556
	Post-test	2.06	0.89		
29. The same instrument can be used for formative purposes and for grading.	Pre-test	3.29	0.46	-2.309	0.021*
	Post-test	3.55	0.51		

\* $p < 0.05$ .

\*\* $p < 0.01$ .

conceptions, ideologies, values, experiences, and knowledge of those who make it" (p. 8). In this sense, subjectivity is unavoidable and inevitably influences important decisions, such as the choice of questions, in the way it is observed, and the proofreading criteria adopted (Fernandes, 2021b).

The significant decrease in the mean of statements 14 and 15 indicates an awareness of the reductive character of the grading, in fact, although it can be considered a form of feedback – evaluative feedback – which expresses judgments about the value or merit of the students' achievements (Tunstall and Gipps, 1996; Gibbs, 2003), it

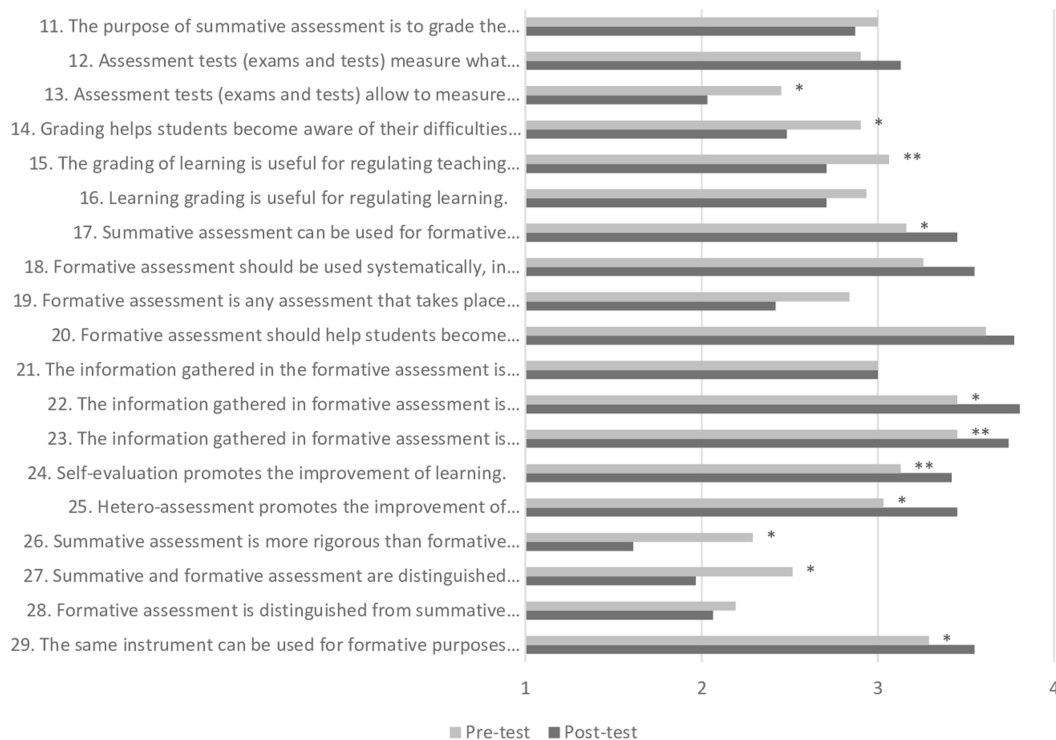


FIGURE 2

Conceptions of summative assessment and formative assessment. 1 = strongly disagree; 2 = disagree; 3 = agree; 4 = strongly agree. \* $p < 0.05$  and \*\* $p < 0.01$ .

does not make the learning and difficulties evidenced by them explicit, and therefore has no formative potential. In fact, “one of the main challenges to be faced in the issue concerning the grading and assigning grades is precisely to think them through so that they will have the pedagogical value that they should have, especially when it comes to supporting students to learn” (Fernandes, 2021a, p. 4). It is noteworthy the fact that several researches refer the need for the development of grading policies that are more suitable to the processes of improving students’ learning and for the regulation of teaching practices (Fernandes, 2021a).

In statements 26 (Summative assessment is more rigorous than formative assessment) and 27 (Summative and formative assessment are distinguished by the instruments used to gather information), the average disagreement increased and was between the spectrum of strongly disagree and disagree. Since these statements contradict the current principles of pedagogical assessment, it was expected that after the course teachers would disagree more with them. What has actually been demonstrated, revealing that they understood that, despite having distinct natures and purposes, both formative and summative assessment are complementary processes that, in order to contribute to the development of student learning, must be rigorous in the processes that involve information gathering (Fernandes, 2019b). It is important to note that there is no evidence or knowledge to support this major misunderstanding regarding this differentiation between the two types of assessment (Fernandes, 2021b).

Teachers also recognized that instruments alone do not distinguish between formative or summative, but the decision-making process should be considered for the differentiation of formative or summative assessment (Popham, 2017). In this regard, any instrument

for information collection can be used in formative or summative assessment practices, “what is really different is the use that is made of the results obtained” (Fernandes, 2019b, p. 7).

It is interesting to note that before the course, the average agreement of teachers with pedagogical training was statistically higher for the statements 17 (Summative assessment can be used for formative purposes), 22 (The information gathered in formative assessment is useful for regulating teaching practices) 23 (The information gathered in formative assessment is useful for regulating learning), 25 (Hetero-assessment promotes the improvement of learning), and 29 (The same instrument can be used for formative purposes and for grading; Table 4). Despite this, it is observed that, with the exception of statements 17 and 25, the mean scores of both groups for the other statements were between agree and strongly agree.

In the post-test, the mean scores of agreement for statements 22 (The information gathered in formative assessment is useful for regulating teaching practices), 23 (The information gathered in formative assessment is useful for regulating learning) and 29 (The same instrument can be used for formative purposes and for grading) remained in the interval between agree and strongly agree, as did the mean score for statement 25 (Hetero-assessment promotes the improvement of learning), which moved into the same interval. However, only for teachers who had not participated in any pedagogical training, the mean scores of the post-test were significantly higher than those of the pre-test. These results demonstrate an evolution in the approach to feedback, being regarded as “a powerful form of communication in the regulation of teaching and learning, as it aims to promote the understanding of the situation

TABLE 4 Differences between the means of the conceptions of teachers with and without pedagogical training on summative and formative assessment in the pre-test ( $n = 31$ ).

Statement	Pedagogical training	N	Mean	Std. deviation	Mann–Whitney U	$p$
11	Yes	10	2.00	0.82	105.000	1.000
	No	21	2.76	0.54		
12	Yes	10	3.60	0.70	103.000	0.926
	No	21	3.24	0.70		
13	Yes	10	3.90	0.32	63.500	0.048*
	No	21	3.67	0.48		
14	Yes	10	2.90	1.20	101.000	0.838
	No	21	3.14	0.65		
15	Yes	10	2.50	1.27	99.500	0.736
	No	21	3.24	0.62		
16	Yes	10	3.50	0.85	86.500	0.285
	No	21	3.33	0.73		
17	Yes	10	3.80	0.42	33.500	0.000***
	No	21	3.19	0.68		
18	Yes	10	3.70	0.48	67.000	0.071
	No	21	3.67	0.48		
19	Yes	10	3.30	0.82	98.000	0.749
	No	21	3.24	0.62		
20	Yes	10	2.80	0.92	76.000	0.147
	No	21	2.67	0.73		
21	Yes	10	3.00	0.67	89.500	0.480
	No	21	3.00	0.45		
22	Yes	10	2.90	0.88	63.000	0.044*
	No	21	2.90	0.62		
23	Yes	10	2.10	0.74	51.000	0.008**
	No	21	2.62	0.50		
24	Yes	10	2.90	0.74	74.500	0.157
	No	21	2.90	0.62		
25	Yes	10	3.10	0.57	54.000	0.017*
	No	21	3.05	0.38		
26	Yes	10	3.10	0.57	72.000	0.106
	No	21	2.86	0.57		
27	Yes	10	3.70	0.48	41.000	0.003**
	No	21	2.90	0.54		
28	Yes	10	3.50	0.85	101.500	0.867
	No	21	3.14	0.48		
29	Yes	10	2.80	0.79	41.500	0.001**
	No	21	2.86	0.79		

\* $p < 0.05$ .

\*\* $p < 0.01$ .

\*\*\* $p < 0.001$ .

for an effective action from the student's and teacher's point of view" (Fialho et al., 2020, p.70).

In the case of statement 17 (Summative assessment can be used for formative purposes), the mean of trained teachers in the pre-test

was 3.70 (SD = 0.48) and that of untrained teachers was 2.90 (SD = 0.54). After the course, the mean of trained teachers remained the same, while the mean of untrained teachers became 3.33 (SD = 0.66), statistically higher than the pre-test. The agreement in the

post-test indicates that, after the course, teachers understood that summative assessment can indeed assume a very relevant formative role in the student learning process (Popham, 2017; Brookhart, 2020) and recognize its regulatory function. Furthermore, it is currently considered that quality summative assessment is subordinate to the principles, methods, and content of formative assessment (Fernandes, 2021b).

In statement 25 (Hetero-assessment promotes the improvement of learning), the mean of teachers with pedagogical training in the pre-test was 3.40 (SD = 0.97) and of those without training was 2.86 (SD = 0.57). In the post-test, the mean scores became 3.60 (SD = 0.52) and 3.38 (SD = 0.50) respectively, with only the mean score of the second group of teachers being significantly higher than the pre-test. Assuming the formative character of hetero-assessment, these results, especially in the group without pedagogical training, point to a greater recognition of the relevance of feedback provided by third parties. In fact, when feedback is provided by peers (students) following well-defined criteria, it is able to promote the improvement of learning for both the evaluated and the evaluators. For Gielen et al. (2011), the use of hetero-assessment, besides complementing the teacher's assessment, allows students to have a greater involvement and motivation in the tasks, making the assessment a secondary objective compared to the task. In addition, it makes assessment a learning method, developing assessment skills in students, who learn how to assess, constituting a tool for active participation in the education path.

In contrast, before the course, the mean agreement of teachers who had no pedagogical training was statistically higher for statements 13 (Assessment tests (exams and tests) allow to measure learning objectively) and 27 (Summative and formative assessment are distinguished by the instruments used to gather information; Table 4). For statement 13, the mean scores of the two groups were between disagree and agree, with the mean scores of teachers with pedagogical training closer to disagree and those of untrained teachers closer to agree. In the post-test, the mean of the teachers with pedagogical training increased slightly, staying on the same spectrum of the scale, while that of the untrained teachers decreased significantly, moving into the range between the points of strongly disagree and disagree.

Teachers' dubiousness about statement 13 reinforces those found in teachers' conceptions of statements 4 (The credibility of the assessment depends on the instruments) and 5 (The quality of the instruments guarantees the objectivity of the assessment), discussed earlier. No assessment instrument is capable of measuring student learning in a completely objective way (Fernandes, 2021b). However, subjectivity can be diminished if the criteria, purposes, and procedures of assessment are well designed and available to all stakeholders in the assessment process, and this practice is not restricted only to the more formal assessment instruments, such as tests and exams, but to any instrument or technique that is used in assessment.

As for statement 27 (Summative and formative assessment are distinguished by the instruments used to gather information), in the pre-test, the mean of trained teachers was exactly at the disagree point, while the mean of untrained teachers was close to agree. In the post-test, although the mean of both groups decreased to 1.70 (SD = 1.06) and 2.10 (SD = 0.99), respectively, only the difference of untrained teachers was significant. This fact reinforces the idea that, after the course, teachers seem to have better understood that it is not the

instruments used that define the formative or summative approach to assessment, but rather, the decisions made from the information gathered.

When the mean scores of conceptions of formative and summative assessment were analyzed between the pre and post-test groups, significant differences were found in 11 out of 19 statements (Table 5), only for the group of teachers who, until then, had not taken any training. Four of these statements dealt only with conceptions related to summative assessment – 13, 14, 15, and 17 –, five dealt with formative assessment – 18, 19, 22, 23, and 25 – and three referred to both conceptions – 26, 27, and 29.

As with the previous group of statements, all the mean scores of the teachers without pedagogical training, in the post-test, which showed significant differences when compared to the pre-test mean scores, demonstrate that the course had a positive influence on their conceptions, especially in this group of teachers. This is because, in the post-test, the means of agreement with the statements show positive changes, that is, an approximation of the meanings that formative and summative assessment assume nowadays.

However, the differences between the averages of statements 13, 15, 17, and 26 stand out, because they are those in which the average agreement in the post-test was located in a different spectrum from the pre-test scale. The differences found between the averages in statements 13 and 17 have already been discussed above. In the case of statement 15 (The grading of learning is useful for regulating teaching practices), the mean of teachers without pedagogical training went from 2.90 (SD = 1.10) to 2.62 (SD = 0.59). A higher rate of disagreement was expected, since grading has little pedagogical value but can be used by teachers to regulate teaching. Nevertheless, it seems necessary to reflect on grading in order to “redefine the purposes and the real meaning of grades” and “relate them closely to the reference learning provided in the curriculum” (Fernandes, 2021a, p. 4).

As for statement 26 (Summative assessment is more rigorous than formative assessment), the mean for this group of teachers decreased from 2.43 (SD = 0.60) to 1.71 (SD = 0.64). This reinforces the fact, discussed earlier, of the contribution of the course in deconstructing the idea that summative assessment is more rigorous than formative assessment.

### 3.3. Conceptions of feedback

Regarding conceptions of feedback, statistical differences were found between teachers' mean agreement in three of the five statements (Table 6; Figure 3). In two of them, statements 31 and 33, the mean in the pre-test was significantly higher than in the post-test and in statement 32 the mean in the post-test was higher.

As in the previous categories, the observed changes were expected, and were in line with the updated literature on feedback in assessment. In statement 32 (The quality of feedback has influence on learning), the mean in the pre-test was already positioned between the points agree and strongly agree, and, in the post-test, it moved closer to strongly agree. In the case of statements 31 (All feedback is useful for learning of students) and 33 (Feedback must be provided to all students in the same way), both the means in the pre-test and post-test were between disagree and agree, with the means in the post-test closer to the disagree point.

TABLE 5 Differences between the means of the conceptions of teachers with and without pedagogical training about summative assessment and formative assessment in the pre and post-test (n = 31).

Statement	Pedagogical training	Test	Mean	Std. deviation	Z	p
11	Yes	Pre-test	3.00	0.67	-0.302	0.763
		Post-test	2.90	0.99		
	No	Pre-test	3.00	0.45	-0.775	0.439
		Post-test	2.86	0.79		
12	Yes	Pre-test	2.90	0.88	-0.378	0.705
		Post-test	3.00	0.82		
	No	Pre-test	2.90	0.62	-1.321	0.186
		Post-test	3.19	0.81		
13	Yes	Pre-test	2.10	0.74	-0.707	0.480
		Post-test	2.30	0.68		
	No	Pre-test	2.62	0.50	-3.419	0.001**
		Post-test	1.90	0.63		
14	Yes	Pre-test	2.90	0.74	-1.134	0.257
		Post-test	2.60	1.07		
	No	Pre-test	2.90	0.62	-1.996	0.046*
		Post-test	2.43	0.68		
15	Yes	Pre-test	3.10	0.57	-0.707	0.480
		Post-test	2.90	1.10		
	No	Pre-test	3.05	0.38	-3.000	0.003**
		Post-test	2.62	0.59		
16	Yes	Pre-test	3.10	0.57	-1.190	0.234
		Post-test	2.70	1.16		
	No	Pre-test	2.86	0.57	-0.647	0.518
		Post-test	2.71	0.56		
17	Yes	Pre-test	3.70	0.48	0.000	1.000
		Post-test	3.70	0.48		
	No	Pre-test	2.90	0.54	-2.179	0.029*
		Post-test	3.33	0.66		
18	Yes	Pre-test	3.50	0.85	0.000	1.000
		Post-test	3.50	0.53		
	No	Pre-test	3.14	0.48	-2.324	0.020*
		Post-test	3.57	0.60		
19	Yes	Pre-test	2.80	0.79	-1.089	0.276
		Post-test	2.30	1.16		
	No	Pre-test	2.86	0.79	-1.615	0.106
		Post-test	2.48	0.93		
20	Yes	Pre-test	3.80	0.42	-1.414	0.157
		Post-test	4.00	0.00		
	No	Pre-test	3.52	0.51	-1.134	0.257
		Post-test	3.67	0.48		
21	Yes	Pre-test	2.80	1.14	-1.000	0.317
		Post-test	3.00	0.94		
	No	Pre-test	3.10	0.70	-0.707	0.480
		Post-test	3.00	0.78		

(Continued)

TABLE 5 (Continued)

Statement	Pedagogical training	Test	Mean	Std. deviation	Z	p
22	Yes	Pre-test	3.70	0.67	-1.000	0.317
		Post-test	3.90	0.32		
	No	Pre-test	3.33	0.48	-2.714	0.007**
		Post-test	3.76	0.44		
23	Yes	Pre-test	3.80	0.42	-1.000	0.317
		Post-test	3.90	0.32		
	No	Pre-test	3.29	0.46	-2.309	0.021*
		Post-test	3.67	0.58		
25	Yes	Pre-test	3.40	0.97	-0.816	0.414
		Post-test	3.60	0.52		
	No	Pre-test	2.86	0.57	-2.840	0.005**
		Post-test	3.38	0.50		
26	Yes	Pre-test	2.00	0.67	-1.730	0.084
		Post-test	1.40	0.52		
	No	Pre-test	2.43	0.60	-2.714	0.007**
		Post-test	1.71	0.64		
27	Yes	Pre-test	2.00	0.47	-1.134	0.257
		Post-test	1.70	1.06		
	No	Pre-test	2.76	0.62	-2.707	0.007**
		Post-test	2.10	1.00		
28	Yes	Pre-test	2.20	0.92	-0.632	0.527
		Post-test	2.00	0.82		
	No	Pre-test	2.19	0.60	-0.361	0.718
		Post-test	2.10	0.94		
29	Yes	Pre-test	3.70	0.48	-0.577	0.564
		Post-test	3.60	0.52		
	No	Pre-test	3.10	0.30	-3.000	0.003**
		Post-test	3.52	0.51		

\*p < 0.05.  
\*\*p < 0.01.

Current literature recognizes that not all feedback is useful for student learning. [Wiliam \(2011\)](#), reviewing the evidence from research regarding feedback over the years, states that poorly designed feedback can be detrimental to student learning. According to the author, in about 40% of the empirical studies, feedback has a negative effect on learning. Conversely, quality feedback, assertively performed, has a positive influence on student learning ([Brookhart, 2008](#)). For the researcher, good feedback: is clear and specific; takes into account aspects related to timing, amount, mode, and audience; describes the work done rather than judging; is positive and suggests improvements; and gives students the help they need to be active subjects in their learning process.

When comparing the mean scores of agreement of the teachers, in the pre-test, considering the variable pedagogical training, it was observed that teachers who had not attended any training obtained the highest mean on statement 33 (Feedback must be provided to all students in the same way;  $U = 36.000$ ;  $p < 0.01$ ). The mean

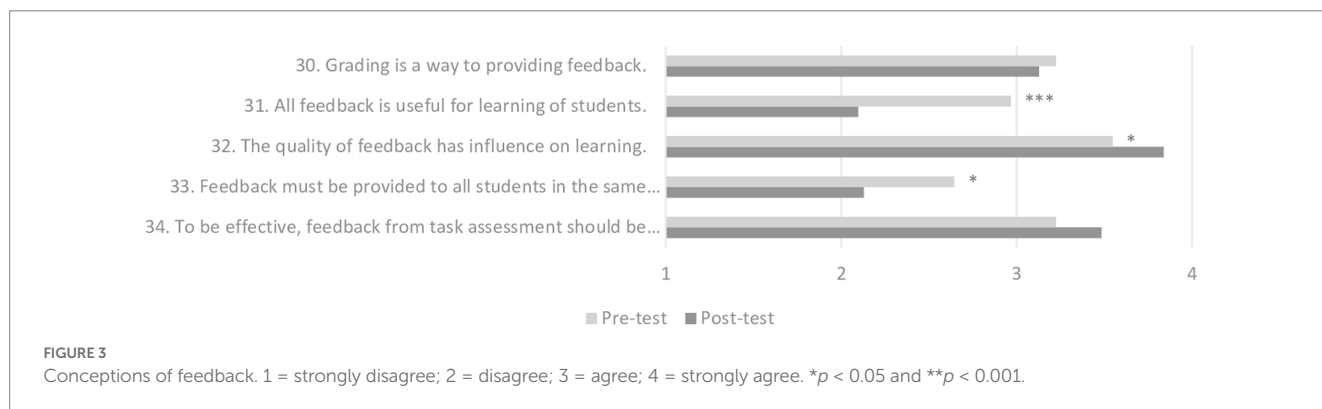
agreement of the trained teachers was 1.90 (SD = 0.88) and that of the untrained teachers was 3.00 (SD = 0.77). This fact shows that teachers who had participated in some training had a more assertive conception of the feedback provided to students, while untrained teachers had a conception of feedback based on principles of equality. Associated with the results of the difference between the mean scores of all teachers in the pre and post-test, which were significantly lower in the post-test, that is, closer to the disagree point of the scale, it can be inferred that participation in the course allowed most teachers to change their conceptions about this aspect of assessment.

Corroborating the results of statement 33, the analysis of the mean of the intragroup conceptions, between the pre and post-test, showed that there was a statistical difference only between the means of the teachers who had not participated in any pedagogical training ([Table 7](#)). The mean in the pre-test was 3.00 (SD = 0.77) and in the post-test was 2.19 (SD = 0.87).

TABLE 6 Differences between the means of all teachers' conceptions of feedback in the pre and post-test (n = 31).

Statement	Test	Mean	Std. deviation	Z	p
30. Grading is a way to providing feedback.	Pre-test	3.23	0.56	-0.832	0.405
	Post-test	3.13	0.62		
31. All feedback is useful for learning of students.	Pre-test	2.97	0.84	-3.488	0.000***
	Post-test	2.10	0.94		
32. The quality of feedback has influence on learning.	Pre-test	3.55	0.51	-2.496	0.013*
	Post-test	3.84	0.37		
33. Feedback must be provided to all students in the same way.	Pre-test	2.65	0.95	-2.447	0.014*
	Post-test	2.13	0.81		
34. To be effective, feedback from task assessment should be given in a short period of time.	Pre-test	3.23	0.67	-1.890	0.059
	Post-test	3.48	0.57		

\*p < 0.05.  
 \*\*\*p < 0.001.



Statistical differences were also found between the means in the pre and post-test, only in the group of teachers without pedagogical training, in the statements 31(All feedback is useful for learning of students), 32 (The quality of feedback has influence on learning), and 34 (To be effective, feedback from task assessment should be given in a short period of time; Table 7). For statements 32 and 34, the means in the post-test, although they approached the strongly agree point, remained between agree and strongly agree. As for statement 31, the average that was between agree and strongly agree in the pre-test, became between disagree and agree in the post-test. For the group of teachers with pedagogical training, the mean agreement in the post-test for statement 31 was also significantly lower than in the pre-test, in the pre-test the mean was 2.60 (SD = 1.17) and in the post-test it was 1.60 (0.70), bringing the post-test mean closer to the strongly disagree point.

Once again, the results have reinforced the importance of the course, in this case, evidenced by the reconceptualization of the concept of feedback, specifically with regard to its usefulness and quality.

### 3.4. Conceptions of grading of learning

Finally, regarding the conceptions of grading of learning, a statistical difference was identified between the mean agreement of all

teachers in only one of the three statements, statement 36 (The final grade must be obtained through a mathematical formula that includes all the grades obtained by the students; Table 8; Figure 4). In this statement, the mean on the post-test (M = 2.45; SD = 0.89) was significantly lower than on the pre-test (M = 3.10; SD = 0.91), demonstrating a decrease in agreement among the teachers.

The results also revealed that the mean agreement of teachers without pedagogical training, for statement 36 (The final grade must be obtained through a mathematical formula that includes all the grades obtained by the students) was also statistically lower in the post-test (Table 9). In the pre-test the mean for this group of teachers was 3.29 (SD = 0.72) and in the post-test it was 2.48 (SD = 0.87). This result corroborates those related to the difference between the mean scores of agreement on the grading of learning, for all teachers in the pre and post-test, noticing that, in the post-test, the mean score decreased significantly due to the change in the conceptions of teachers without pedagogical training, from 3.29 (SD = 0.72) to 2.48 (SD = 0.87), standing between disagree and agree on the scale. Nevertheless, it is an “indicator of a shift to a more holistic, thoughtful, and reflective perspective of grading, as opposed to a grading that results from a mathematical formula applied in a mechanical way” (Fialho and Cid, 2021, p.166).

When comparing the mean scores of agreement of teachers before the course with regard to the pedagogical training variable, a significant difference is observed only for statement 37 (In the final

TABLE 7 Differences between the means of the conceptions of teachers with and without pedagogical training about feedback in the pre and post-test (n = 31).

Statement	Pedagogical training	Test	Mean	Std. deviation	Z	p
30	Yes	Pre-test	3.20	0.63	-0.577	0.564
		Post-test	3.10	0.57		
	No	Pre-test	3.24	0.54	-0.632	0.527
		Post-test	3.14	0.66		
31	Yes	Pre-test	2.60	1.17	-2.041	0.041*
		Post-test	1.60	0.70		
	No	Pre-test	3.14	0.57	-2.818	0.005**
		Post-test	2.33	0.97		
32	Yes	Pre-test	3.80	0.42	0.000	1.000
		Post-test	3.80	0.42		
	No	Pre-test	3.43	0.51	-2.714	0.007**
		Post-test	3.86	0.36		
33	Yes	Pre-test	1.90	0.88	-0.276	0.783
		Post-test	2.00	0.67		
	No	Pre-test	3.00	0.77	-2.984	0.003**
		Post-test	2.19	0.87		
34	Yes	Pre-test	3.50	0.53	-0.577	0.564
		Post-test	3.40	0.52		
	No	Pre-test	3.10	0.70	-2.310	0.021*
		Post-test	3.52	0.60		

\*p < 0.05.

\*\*p < 0.01.

TABLE 8 Differences between the means of all teachers' conceptions of grading of learning on the pre and post-test (n = 31).

Statement	Test	Mean	Std. deviation	Z	p
35. The final grade should take into account the information gathered in the summative and formative assessment.	Pre-test	3.13	0.76	-1.662	0.096
	Post-test	2.87	0.96		
36. The final grade must be obtained through a mathematical formula that includes all the grades obtained by the students.	Pre-test	3.1	0.91	-2.553	0.011*
	Post-test	2.45	0.89		
37. In the final grading, the assessment tests (exams) should have a higher "weight" than other summative assessment instruments or tasks.	Pre-test	2.16	0.73	-1.806	0.071
	Post-test	1.84	0.58		

\*p < 0.05.

grading, the assessment tests (exams) should have a higher "weight" than other summative assessment instruments or tasks; U = 30.000; p < 0.001). The mean of teachers who had already attended training was 1.50 (SD = 0.53), statistically lower than the mean of those who had no training, which was 2.48 (SD = 0.60), demonstrating that most of the teachers with pedagogical training and less than half of the teachers without this training disagreed, partially or totally, with the attribution of greater importance ("weight") to the assessment tests (tests and exams) in the final classification.

In the post-test, the average agreement of the teachers who had pedagogical training was 1.70 (SD = 0.48) and that of the untrained teachers was 1.90 (SD = 0.62), the latter significantly lower than the

pre-test average, demonstrating the importance of the course in changing their conceptions in this regard. In the post-test, the average agreement of the teachers who had previous training was 1.70 (SD = 0.48) and the average agreement of those who had no training was 1.90 (SD = 0.62), the latter significantly lower than the pre-test average, showing the importance of the course in changing their conceptions in this regard.

The results obtained in statement 37 seem to show that teachers diversify the instruments and tasks of summative assessment, without giving priority to attendance and/or examination, for the purpose of final classification. Although this diversification is positive, it is not guaranteed that the assessment assumes a formative nature, "because

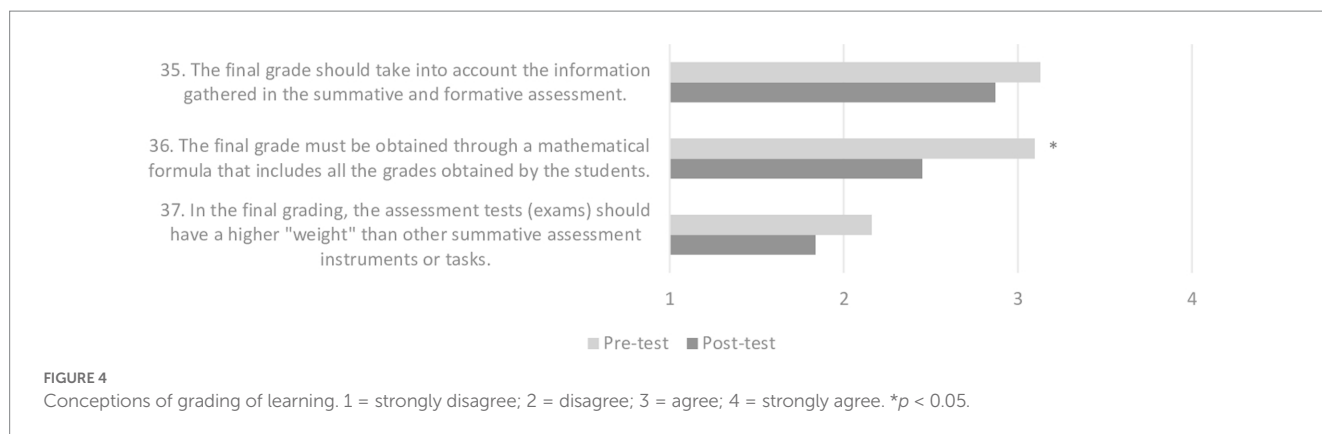


TABLE 9 Differences between the means of the conceptions of teachers with and without pedagogical training about grading of learning in the pre and post-test (n = 31).

Statement	Pedagogical training	Test	Mean	Std. deviation	Z	p
35	Yes	Pre-test	2.80	0.92	-1.134	0.257
		Post-test	2.50	0.97		
	No	Pre-test	3.29	0.64	-1.232	0.218
		Post-test	3.05	0.92		
36	Yes	Pre-test	2.70	1.16	-0.604	0.546
		Post-test	2.40	0.97		
	No	Pre-test	3.29	0.72	-2.773	0.006**
		Post-test	2.48	0.87		
37	Yes	Pre-test	1.50	0.53	-1.000	0.317
		Post-test	1.70	0.48		
	No	Pre-test	2.48	0.60	-2.398	0.016*
		Post-test	1.90	0.63		

\* $p < 0.05$ .

\*\* $p < 0.01$ .

the nature of assessment is not in the instruments and techniques, but in the intentionality with which they are used, that is, in the use that is given to the information collected” (Fialho et al., 2020, p. 79).

## 4. Conclusion

In line with the objectives that guided this study, the results obtained allowed identifying several conceptions of assessment among higher education teachers, as well as changes in conceptions. Regarding the category of conceptions of pedagogical assessment, the change in conceptions of the objectives of pedagogical assessment was mainly highlighted, overcoming the misconception that the main objective is the grading of learning. However, it was also evident that there is a need for further clarification on the aspects that address the subjectivity/objectivity of assessment.

In the case of the category of conceptions of formative and summative assessment, significant changes in conceptions were observed, mainly in the group of teachers without pedagogical training. Changes in conceptions regarding the assessment instruments are highlighted, whose intentionality in use determines

the formative or summative character, since the nature of the assessment is not in the instruments and techniques, but in the use given to the information gathered. It also highlights the understanding that both formative and summative assessment demand rigor and credibility and that summative assessment may, on certain occasions, be used for formative purposes. It was noted, however, that the rate of disagreement with regard to the classification of learning fell short of what was expected, since, although it may be useful for regulating teaching, the classification has little pedagogical value.

As for the category of conceptions of feedback, the results showed the importance of the course for a change in conceptions, particularly regarding its usefulness and quality. The fact that teachers, especially those who did not have pedagogical training, understood that not all feedback is useful for learning and, if poorly planned and used, it may even be harmful, should be highlighted.

Finally, in the category of conceptions of the grading of learning, it was evident that, in general, some of the teachers’ previous conceptions were already in line with the current literature on pedagogical assessment and that the idea that the final grade should be composed of the sum of all grades obtained by the students throughout the course seems to have been overcome.

The data gathered also made it possible to assess the impact of the course, whose importance for the pedagogical training of higher education teaching staff was well evidenced in this study. The comparison of results between teachers with and without pedagogical training showed that the former's conceptions of assessment are more consistent with contemporary trends than those of the latter. On the other hand, the positive impact of the Training Course on Assessment in Higher Education was also very clear, since the comparison of the results of the pre-test with the post-test shows that relevant changes have occurred towards the most current theories of pedagogical assessment, these changes being more significant in the group of teachers without any pedagogical training. However, even though some more traditional conceptions persist, which is not surprising because conceptual change is a complex process that requires time and persistence, the need to invest in the continuous training of teachers is reinforced.

Higher education institutions must respond to the Bologna challenges and to the demands of quality education (4 SDG), countering the strongly entrenched instructional teaching culture. This response requires pedagogical qualification policies for teachers, within the institutional framework of university autonomy, which promote the development of pedagogical skills that generate changes and innovation in teaching practices, leading to a new way of being and being a teacher in higher education.

The study presented a limitation in the number of teachers surveyed, which corresponded to the number of teachers who participated in the three editions of the Grounding and Improving Pedagogical Assessment in Higher Education course. Although the results are not generalizable to the population of university teachers of the participating institution or of Portugal, there is important information that can be added to the results of other studies, allowing for a greater reflection on the conceptions and practices of pedagogical evaluation of higher education teachers.

However, this exploratory study contributed to the recognition of the strategic importance of pedagogical training, creating opportunities for the validation questionnaire, thus, new editions of the course will allow these to continue to be applied in order to obtain representative data regarding the conceptions and practices of pedagogical evaluation of higher education teachers.

## References

- Alcántara-Rubio, L., Valderrama-Hernández, R., Solís-Espallargas, C., and Ruiz-Morales, J. (2022). The implementation of the SDGs in universities: A systematic review. *Environ. Educ. Res.* 28, 1585–1615. doi: 10.1080/13504622.2022.2063798
- Aleixo, A. M., Azeiteiro, U., and Leal, S. (2019). A implementação dos objetivos do desenvolvimento sustentável no ensino superior português, 1.ª Conferência Campus Sustentável, Porto, Portugal.
- Almeida, M. M. (2020). Formação pedagógica e desenvolvimento profissional no ensino superior: perspectivas de docentes. *Rev. Brasil. Educ.* 25, 1–22. doi: 10.1590/s1413-24782019250008
- Barreira, C., Bidarra, G., Monteiro, F., Vaz-Rebello, P., and Alferes, V. (2017). Avaliação das aprendizagens no ensino superior. Percepções de professores e estudantes nas universidades portuguesas. *Rev. Iberoamericana Educ. Superior* 8, 24–36.
- Black, P. (2013). "Pedagogy in theory and in practice: formative and summative assessments in classrooms and in systems" in *Valuing assessment in science education: Pedagogy, curriculum, policy*. eds. D. Corrigan, R. Gunstone and A. Jones (New York: Springer), 207–229.
- Black, P. J., Harrison, C., Lee, C., Marshall, B., and Wiliam, D. (2011). *Assessment for learning. Putting it into practice*. Maidenhead, U.K.: Open University Press.
- Brookhart, S. M. (2008). *How to give effective feedback to your students*. Alexandria: ASCD.
- Brookhart, S. M. (2020). "Feedback and measurement" in *Classroom assessment and educational measurement*. eds. S. M. Brookhart and J. H. McMillan (New York: Routledge), 63–78.
- Chaleta, E., Saraiva, M., Leal, F., Fialho, I., and Borralho, A. (2021). Higher education and sustainable development goals (SDG) - contribution of the undergraduate courses of the School of Social Sciences of the University of Évora. *Sustainability* 13:1828, 1–11. doi: 10.3390/su13041828
- ESG (2015). Standards and Guidelines for Quality Assurance in the European Higher Education Area. Brussels, Belgium.
- European Commission (2010). Improving teacher quality: the EU agenda. European Commission. Available at: [http://www.mv.helsinki.fi/home/hmniemi/EN\\_Improve\\_Teacher\\_Quality\\_eu\\_agenda\\_04\\_2010\\_EN](http://www.mv.helsinki.fi/home/hmniemi/EN_Improve_Teacher_Quality_eu_agenda_04_2010_EN) (Accessed December 23, 2022).
- Federação Académica do Porto (FAP) (2021). Inovação pedagógica. Ventos de mudança no ensino superior. FAP.
- Fernandes, D. (2004). *Avaliação das aprendizagens: Uma agenda, muitos desafios*. Cacém: Texto Editores.
- Fernandes, D. (2019a). Avaliação formativa. Folha de apoio à formação – Projeto MAIA. Instituto de Educação da Universidade de Lisboa e Direção Geral de Educação do Ministério da Educação.
- Fernandes, D. (2019b). Avaliação sumativa. Folha de apoio à formação – Projeto MAIA. Instituto de Educação da Universidade de Lisboa e Direção Geral de Educação do Ministério da Educação.
- Fernandes, D. (2021a). Avaliação Pedagógica, Classificação e Notas: Perspetivas Contemporâneas. Folha de apoio à formação - Projeto de Monitorização, Acompanhamento e Investigação em Avaliação Pedagógica (MAIA). Ministério da Educação/Direção-Geral da Educação.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

IF and MCI contributed to conception and design of the study. MCo performed the statistical analysis. IF, MCI, and MCo wrote sections of the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was financed by national funds from the Foundation for Science and Technology (FCT), I.P., within the scope of the project UIDB/04312/2020, and the Ph.D. Research Grant with reference UI/BD/151034/2021.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Fernandes, D. (2021b). Para uma fundamentação e melhoria das práticas de avaliação pedagógica no âmbito do Projeto MAIA. Texto de Apoio à formação - Projeto de Monitorização Acompanhamento e Investigação em Avaliação Pedagógica (MAIA). Ministério da Educação/Direção-Geral da Educação.
- Fialho, I., Chaleta, E., and Borralho, A. (2020). Práticas de avaliação formativa e feedback, no ensino superior. In M. Cid, N. Rajadell Puiggròs and G. Santos Costa (Coords.), *Ensinar, avaliar e aprender no ensino superior: perspetivas internacionais*. (Évora: CIEP-UE), 65–92.
- Fialho, I., and Cid, M. (2021). “Fundamentar e melhorar a avaliação pedagógica no ensino superior. Um processo formativo sustentado na investigação-ação em contexto digital” in *Portas que o digital abriu na investigação em educação*. eds. A. Nobre, A. Mouraz and M. Duarte (Lisboa: Universidade Aberta), 151–173.
- Field, A. (2009). *Discovering statistics using SPSS*. 3rd Edn. Thousand Oaks: SAGE Publications, Inc.
- Flores, M. A., Carvalho, A., Arriaga, C., Aguiar, C., Alves, F., Viseu, F., et al. (2007). *Perspectivas e estratégias de formação de docentes do ensino superior. Relatório de investigação*. Braga: CIED.
- Flores, M. A., Veiga Simão, A. M., Barros, A., Flores, P., Pereira, D., Fernandes, E., et al. (2021). Ensino e aprendizagem à distância em tempos de COVID-19. Um estudo com alunos do Ensino Superior. *Rev. Portug. Pedagogia* 55, e055001–e055028. doi: 10.14195/1647-8614\_55\_1
- Ghiglione, R., and Matalon, B. (1992). O inquérito. Teoria e prática. Celta Editora.
- Gibbs, G. (2003). “Uso estratégico de la evaluación en el aprendizaje” in *Evaluar en la universidad. Problemas y nuevos enfoques*. ed. S. G. Brown (Madrid: Narcea), 61–75.
- Gielen, S., Dochy, F., Onghena, P., Struyven, K., and Smeets, S. (2011). Goals of peer assessment and their associated quality concepts. *Stud. High. Educ.* 36, 719–735. doi: 10.1080/03075071003759037
- Hill, M. M., and Hill, A. (2005). *Investigação por questionário*. Lisboa: Edições Sílabo.
- Kemmis, S., and McTaggart, R. (1992). *Como planificar la investigación-acción*. Barcelona: Editorial Laertes.
- Ministério da Educação e Ciência (2015). *Experiências de inovação didática no ensino superior*. Lisboa: MEC.
- Nitko, A. J., and Brookhart, S. M. (2014). *Educational assessment of students*. 6th Edn. Edinburgh: Pearson.
- Ogawa, M., and Vosgerau, D. (2019). Formação docente do ensino superior: o papel das instituições. *Rev. Espacios* 40, 1–7.
- Pereira, F. S., and Leite, C. (2020). O Processo de Bolonha na sua relação com a agenda da qualidade – uma análise focada no perfil dos docentes que asseguram os cursos de educação básica. *TMQ – Techniques, Methodologies and Quality Número Especial – Processo de Bolonha*, 135–150.
- Pereira, A., Oliveira, I., Tinoca, L., Pinto, M. C., and Amante, L. (2015). *Desafios da Avaliação Digital no Ensino Superior*. Lisboa: Universidade Aberta.
- Popham, W. J. (2017). *Classroom assessment: What teachers need to know*. 8th Edn. Los Angeles: Pearson.
- Russell, M. K., and Airasian, P. W. (2008). *Classroom assessment: Concepts and applications*. New York: McGrall-Hill.
- Salvia, A. L., Leal Filho, W., Brandli, L. L., and Griebeler, J. S. (2019). Assessing research trends related to sustainable development goals: local and global issues. *J. Clean. Prod.* 208, 841–849. doi: 10.1016/j.jclepro.2018.09.242
- Tunstall, P., and Gipps, C. (1996). Teacher feedback to young children in formative assessment: a typology. *Br. Educ.* 22, 389–404. doi: 10.1080/0141192960220402
- UNESCO (2017). *Educação para os Objetivos de Desenvolvimento Sustentável. Objetivos de aprendizagem*. Paris: Unesco.
- Wiliam, D. (2011). *Embedded formative assessment*. Bloomington: Solution Tree Press.
- Wiliam, D. (2017). Assessment and learning: some reflections. *Assess. Educ.* 24, 394–403. doi: 10.1080/0969594X.2017.1318108
- Xavier, A. R. C., and Leite, C. (2019). Mapeamento da formação pedagógica de docentes universitários nas universidades públicas portuguesas. *Rev. Lusófona Educa.* 45, 109–123. doi: 10.24140/issn.1645-7250.rle45.08