

# Observing Classes at Técnico (2010-2019): Do observations impact on the quality of teaching?

Filipa David  
Statistics and Prospective Unit  
(IST, ULisboa)  
Lisboa, Portugal  
filipasdavid@tecnico.ulisboa.pt

Leonor Moura  
Academic Development Unit  
(IST, ULisboa)  
Lisboa, Portugal  
leonor.moura@tecnico.ulisboa.pt

Patrícia Simões  
Academic Development Unit  
(IST, ULisboa)  
Lisboa, Portugal  
patriciamsimoes@tecnico.ulisboa.pt

Gonçalo Moura  
Academic Development Unit  
(IST, ULisboa)  
Lisboa, Portugal  
goncalo.moura@tecnico.ulisboa.pt

Marta Graça  
Academic Development Unit  
(IST, ULisboa)  
Lisboa, Portugal  
marta.graca@tecnico.ulisboa.pt

Isabel Gonçalves  
Academic Development Unit  
(IST, ULisboa)  
Lisboa, Portugal  
isabel.goncalves@tecnico.ulisboa.pt

**Abstract**—The main objective of the observation of classes project is to promote and improve the pedagogical skills of professors and their professional development. This paper refers to three existing contexts of observation and aims to understand how the class observation process is reflected in the pedagogical performance of teachers over time, using the Quality of Curricular Units (QUC) results in order to provide details about the performance of the teachers in a certain semester. This study focuses on the universe of more than 700 teachers evaluated by QUC, from 2010/2011 to 2019/2020 at Técnico. The statistical analysis demonstrates evidence that there are differences between the pedagogical performance of teachers over the years, namely with an emphasis on moments after the observation of classes has occurred. Qualitative results from the feedback meetings, after every observation process, reveal that class observations are seen as an appropriate means to promote the quality of teaching and the pedagogical development of teachers. More work needs to be done in this area, on the one hand to continue monitoring the professors' achievements and their performance over the years at Técnico, and on the other hand, to identify other variables that influence the pedagogical skills of professors, assessing and improving the positive impact of the classroom observations done by non-peer professionals.

**Keywords** – classes observation, higher education, feedback, quality assurance, teaching skills

## I. INTRODUCTION

Teaching is characterized by being a challenging and demanding activity. In this sense, the Pedagogical Council (CP), the Academic Development Unit (NDA) and the Statistics and Prospective Unit (NEP) collaborated to implement a class observation system within Instituto Superior Técnico (Técnico).

The main objective of this project is to **promote and improve the teaching skills of teachers**, promoting their professional development. This is supported by the literature, as according to Reis “there has been an international and national tendency to view the observation of classes as a process of professional interaction, of an essentially formative nature, centered on individual and of teachers and improving the quality of teaching and learning” [1].

Currently, the observations of classes at Técnico take place within the scope of the Shaping the Future Program (PSF -

<http://shapingthefuture.tecnico.ulisboa.pt/>) and the Monitoring of Teachers with Pedagogical Performance to Improve (DDPaM), and eventually, observations of classes are made on request, whether by teachers, departments or others.

Observations are made by one or two professionals from the NDA team and, typically, each teacher will be observed once per school year or, more infrequently, once each semester of the school year.

In order to standardize the criteria for the observations, the NDA team developed an Observation Grid, based on the available literature. This Grid is previously made available to Professors, so that they can learn about the aspects that will be considered during the observation. This grid is available even for teachers who are not currently being observed.

Regarding the scheduling of observations, within the scope of the PSF and DDPaM Programs, NDA makes a first contact, at the beginning of each academic year, in order to ascertain the availability of teachers. The class chosen for observation in each semester/academic year, is identified by the teacher, and it can be a practical, theoretical, tutorial or laboratory class. The date and timing of the observation are also chosen by the teacher. In the case of observations by request, the teacher takes the initiative to contact the NDA team, showing their interest and availability for the activity. Usually teachers will choose classes they want to improve, or were students' feedback in former academic years has been critical.

After the observation, a meeting is held between the teacher and the member(s) of the NDA team, with the aim of jointly reflecting on the observed class and providing feedback on the positive aspects and the potential improvements to be implemented in the classes and/or in the teacher's pedagogical practices. This feedback, as well as the filled grid of observation are then sent to the teacher in a PDF format, for further reference. Extra support documents can also be provided, on request, or to complement the feedback given. Extra training can also be provided (offers provided by the Training and Development Program), in order to support

aspects of the teaching and learning activities in need of improving.

An analysis of the impact of this observation practices in the pedagogical performance of IST's teachers will be presented, within the scope of the three existing programs (PSF, DDPaM and Other).

In order to ascertain how the Class Observation process is reflected in the pedagogical performance of teachers over time, (Técnico Course Units Quality System) results were considered (<https://tecnico.ulisboa.pt/en/>). The system proves to be a crucial tool for this study, since one of the measures of analysis of this process is precisely the performance of the teachers in a respective semester, for a given type of class. The indicator used is a measure calculated using the dimensions of analysis of the survey itself, coming from students' assessment of the quality of the classes taught at IST [2, 3, 4].

## II. CLASSES OBSERVATION AT IST: HOW IT IS DONE

The Pedagogical and Scientific Councils at IST, following best international practices for improving the quality of teaching and learning at IST, started a Classes' Observation Program associated with the Shaping the Future Program (PSF), which "promotes the integration and adaptation of junior faculty and researchers to IST's culture, through a Mentoring Program and a set of initiatives aiming at accelerating the development of their careers in the dimensions of scientific and academic leadership and the development of competences in scientific and pedagogical areas (training and lectures observation)" (<http://shapingthefuture.tecnico.ulisboa.pt/en/>).

Later, and during the assessments of this Program (PSF), the idea of observing other teachers at Técnico came up, and so another initiative was launched, which was observing teachers assessed by students as needing to improve the quality of their classes, as well as other teachers who felt they might profit from this quality enhancement practice.

The process inherent to the practice of classroom observation is subject to continuous monitoring and evaluation, both by the NDA team, who carry out the observations and by the teachers who are observed. All observations consist of three phases, that correspond to the moments before, during and after the observations. Before any observation, the teacher is contacted and information is collected on the type of class to be observed, in order to guide and prepare the observation, adjusting it to the context, leaving the teacher to choose the class to observe (day, time and type of class).

Before the actual observation, the observation grid is sent to the teacher for his/her knowledge. Some examples of activities observed that are included in this grid are: starts by briefly summarizing the endpoint of the previous class; organizes the class so that the relationship between objectives and activities is clear; speaks perceptibly, with sufficient volume and appropriate speed; has eye contact with students throughout the classroom; proposes exercises and/or practical

activities for students to carry out, providing clear instructions; analyses and discusses with the group the challenges and the results achieved; explicitly shows enthusiasm for the class subject; encourages students' participation; identifies and positively reinforces students' participation; promotes activities that are both reasonable and challenging; Ensures a balanced relationship between the complexity/volume of content transmitted and the time available.

During the observation, the evidences observed in class are recorded on the grid, according to the following areas: class organization; presentation; content; practical activities and classroom climate. There is also room in the grid to make notes on positive aspects of the class and those that can be improved.

The observation grid is a fundamental tool in this activity, being regularly improved based on all the empirical knowledge collected and also in the literature. After the observation, the information collected is harmonized and a feedback meeting is scheduled with the teacher. In this meeting, the observers transmit to the teacher, in a reasoned and dialogued way, the positive, innovative and improving aspects of the observed class and a joint reflection is made about the observed class, outlining strategies aimed at developing the pedagogical skills of the teacher. At this meeting, there is total openness for teachers to comment on the NDA's intervention.

The feedback component after observing the classes is deemed central to the quality of the observation, as commented on by observed teachers. This moment of analysis of the collected data, dialogue and reflection between the observer team and the teacher has six main objectives: the identification of situations that can be improved or maintained; the description of students' and teachers' behaviors during the class; the analysis of the more challenging situations in the context of the class; making suggestions on how to improve (e.g. literature, complementary trainings); encouraging self-assessment and sharing information and knowledge.

The fact that this process is structured and divided into standardized moments facilitates the monitoring and evaluation of the entire process.

The documentation of this good practice, which implied reflection on the whole process, as well as all the statistical work developed by NEP, were two important moments of evaluation of this activity that has been carried out at IST since 2015/2016.

## III. METHODOLOGY

In this analysis, the teachers are identified as having been observed by one of the three distinct programs (PSF, DDPaM or Other), and there may be exceptional cases in which the observation to which they were submitted may result from more than one process. They will then be distinguished by groups A, B and C.

In order to characterize the group of teachers evaluated by the system through the survey of students, assessments of the teachers who participated in the class observation program were singled out. Results of student's assessment between the academic years of 2010/2011 and 2019/2020 were collected, if they were considered representative.

The pedagogical performance of teachers was analyzed according to the RADIST classification, obtained by calculating the arithmetic mean of the median of answers to the Teaching Staff questions' of 'Advantages from in-class learning', 'Pedagogical ability' and 'Interaction with students' groups. This classification is applied in the scope of the IST Faculty Performance Evaluation Regulation.

In order to assess the teaching performance of teachers per academic year, the research team decided to work with only one annual value. To obtain this annual value, the average of the various values of RADIST was considered, weighted with the number of responses given to the teacher-course unit-type of class trio. Similar to what is regulated in the system, the categorization (and also coloring) of this measure was adopted in this study, which takes values between 1 and 9, as follows:

1. [1; 3]: **Inadequate results**
2. ]3; 5]: **Results to improve**
3. [5; 7[: **Regular results**
4. [7; 8[: **Regular results +**
5. [8; 9[: **Very Good results**
6. [9]: **Excellent results**

#### IV. CHARACTERIZATION OF THE TEACHING POPULATION

The universe of Técnico's teachers considered all those that were assessed within the scope of Técnico's QUC system through the survey applied to students, between the years 2010/2011 and 2019/2020, whose results were representative. The characterization of the teacher's universe in study will be presented according to their sex, age, nationality and category.

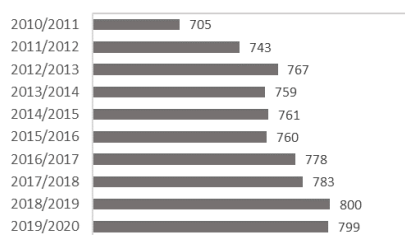


Fig. 1. Universe of teachers evaluated by since 2010/2011.

In the last 10 academic years, more than 700 teachers have been evaluated through surveys of students within the scope of the QUC system, per academic year (Fig.1). From the universe of teachers considered, there is a constant trend of distribution by sex, where the number of male teachers is about 3 times higher than the number of female teachers, which have been around to 26% through the last ten years.

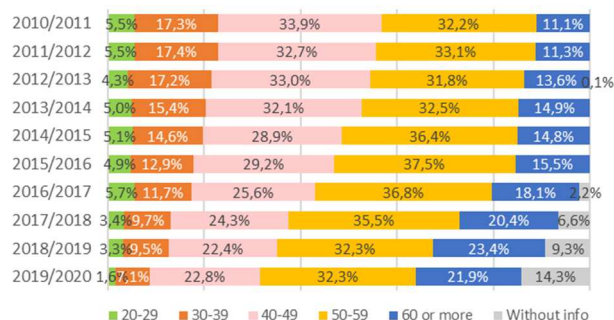


Fig.2. Universe teachers evaluated by the by Age Group (in years).

There is a more evident decreasing trend over the years of the 30 to 39 age group, which in 2010/2011 was 17.3% and in the last year only 7.1%. In the age group of 40 to 49 years, there is also a decrease by about 11% until 2019/2020. The decrease in this age groups is both compensated by the growing trend of the older age group which in 2010/2011 represented 11.1% and in 2019/2020, 21.9%. These data has missing information in the past 4 academic years, regarding the faculty's date of birth (Fig. 3).

Among the universe of teachers under analysis, there is a relatively low percentage of teachers with foreign nationality, around 2% to 3%.

The majority of teachers has the type of category "Professor", followed by those who have another type of category, called "Others". In the "Professor's" category we find mostly Assistant Professors (>50%) or Associate Professors (>20%), with the invited Professors being the least represented. Research professors are underrepresented in the analysis, with only 1% in the last two academic years (Fig. 3), mostly Assistant Researchers.

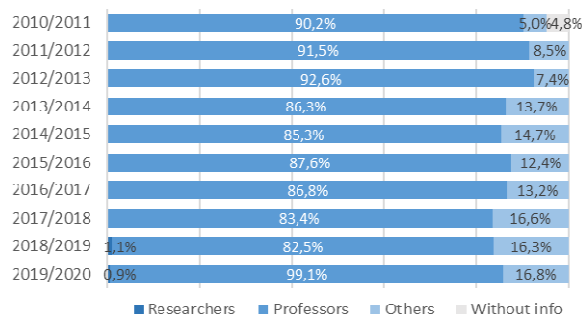


Fig. 3. Universe of teachers evaluated by the QUC by Category Type.

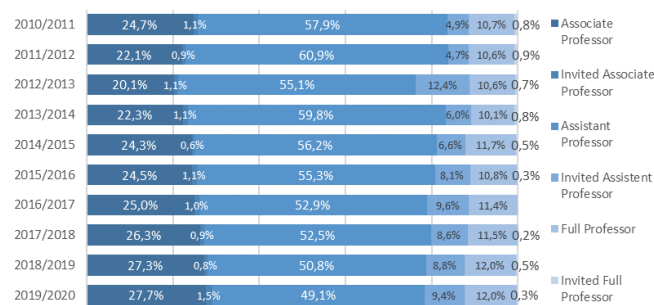


Fig. 4. Universe of teachers evaluated by the QUC by Category of Professor.

Overall, there is a slight downward trend in the proportion of Assistant Professors, which was 57.9% in 2010/11 and 49.1% in 2019/20, and an increase in the proportion of Associate

Professors over the years, as seen also on Full Professors proportion, despite residual over the years (from 10.7% in 2010/11 to 12% in 2019/20).

Based on these numbers, the same information is followed by the classroom observation program (Group A, Group B and Group C). Table 1 shows the number of teachers in analysis who were enrolled in the observation of classes program, within the scope of each of the three programs under study.

TABLE I. UNIVERSE OF TEACHERS WHOSE CLASSES WERE OBSERVED WITHIN THE SCOPE OF EACH PROGRAM (GROUP A, GROUP B AND GROUP C)

Academic Year	Observed Teachers		
	Group A	Group B	Group C
2015/2016	16	-	-
2016/2017	14	25	-
2017/2018	18	2	-
2018/2019	47	9	1
2019/2020	21	1	2

As Table I shows, the observations of group A teachers' classes were carried out in all academic years between since 2015/2016, with more occurrences in 2018/2019. The observations of group B teachers' classes started in 2016/2017, showing a decreasing trend. In 2019/2020, due to the COVID-19 pandemic, it was not possible to make further observations. The observation of classes by teachers in group C occurred less frequently compared to the other groups.

The characterization of the teachers' "profile" who were enrolled in the classes' observation program, in each of the groups in Table I, will be presented.

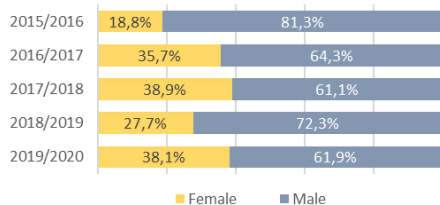


Fig. 5. Proportion of teachers observed in Group A, by Sex.

In the case of teachers observed in group A, although a higher proportion of male teachers observed was identified, it should be noted that in the year with the highest number of observations (2018/2019), only 27.7% were from classes taught by female teachers.

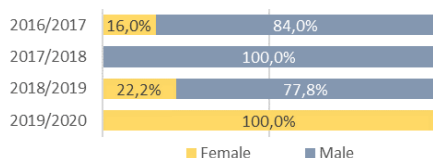


Fig. 6. Proportion of teachers observed in Group B, by Sex.

Fig. 6 reveals that more male teachers than female teachers were observed through the observation years, and in 2019/2020 only one female teacher was observed. In group C, the teachers enrolled in the classes' observation are all male.

Regarding the nationality of teachers observed, in group A, the majority is Portuguese, highlighting the 21.4% foreign

teachers observed in 2016/2017. In group B, the proportion of teachers with foreign nationality was residual in 2016/2017 and non-existent in the following years. In group C, all teachers whose classes were observed are Portuguese.

The distribution of teachers by group of classes' observations according to the age group (with age intervals in years) is shown in the next figures.

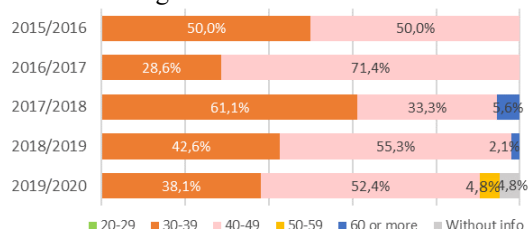


Fig. 7. Proportion of teachers observed in Group A, by Age Group (in years).

In group A, most of the observed teachers were between 40 and 49 years old in almost all years of analysis, except for the year 2017/2018, in which 61.1% were between 30 and 39 years old. Few were the teachers with more than 60 years old observed in this group. In fact, teachers close to retirement usually are not enrolled in the programs A, B and C.

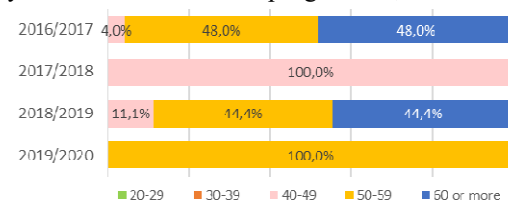


Fig. 8. Proportion of teachers observed in Group B, by Age Group (in years).

In group B, it can be seen that the teachers observed are all over 40 years old, and that in 2016/2017 and 2018/2019, teachers observed were over the age of 50 years (Fig. 8). In group C, all teachers enrolled in class observation were between 54 and 61 years old.

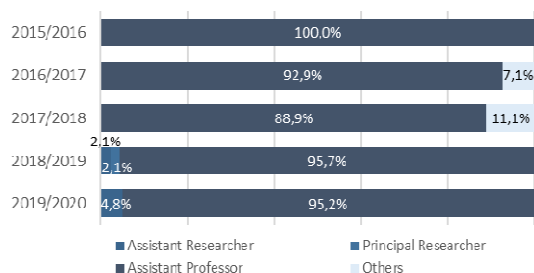


Fig. 9. Proportion of teachers observed in Group A, by Category.

In group A, the teachers observed are mostly Assistant Professors, but we still can highlight the 7.1% and 11.1% of teachers with the category "Others" (which includes categories that are not considered a tenured career) in 2016/2017 and 2017/2018.

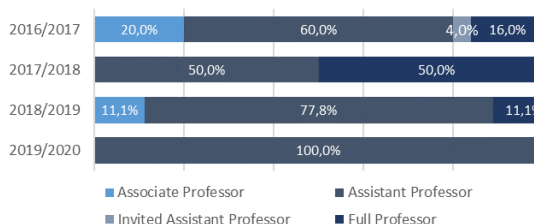


Fig. 10. Proportion of teachers observed in Group B, by Category.

Also, in group B, most of the observed teachers have the category of Assistant Professor or Full Professor (Fig. 10). In group C, each of the observed teachers has one of the categories of tenured career.

### V. CHARACTERIZATION OF TEACHERS' PEDAGOGICAL PERFORMANCE

In order to understand the behavior of the teaching performance of teachers over the academic years under study, the teaching population will be characterized according to sex, age, nationality and teaching category, based on the grouping criteria of the RADIST classification, weighted by the six categories defined previously.

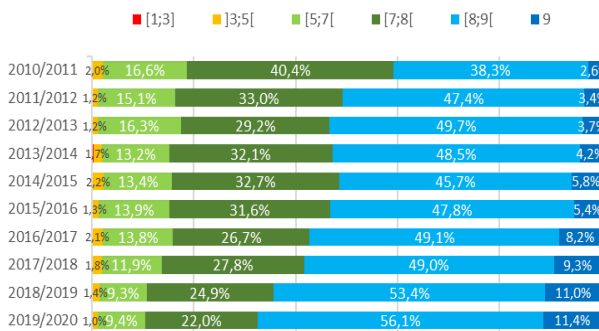


Fig. 11. Pedagogical performance of teachers according to the weighted RADIST classification.

In Fig. 11, we can see an increasing trend of evaluated teachers with “very good” and “excellent” results over the years, in contrast to the decrease of teachers with “regular +” results. Despite being residual, the cases with results “to improve” are less than 2% per academic year. The cases with “inadequate” results are lower than 0.3%. As seen, there is a growing trend of teachers with very good or excellent results.

However, when analyzed by sex, the most evident differences are found in the next two figures (Fig. 12 and 13), for both female and male teachers’ results, based on RADIST classification: the fluctuation in the trend ‘improving towards excellency’, in the case of female teachers notices a decrease of close to 1.8% in the group of teachers with “excellent” and “regular +” results in the year 2012/2013, compared to previous years. In 2013/2014, we find a case with an “inadequate” result (that did not occur in any other year), but the cases with an “excellent” result rises again; in the year 2014/2015 we see a decrease in cases with “very good” results, compared to the previous year, accompanied by an increase in cases with “regular +” results.

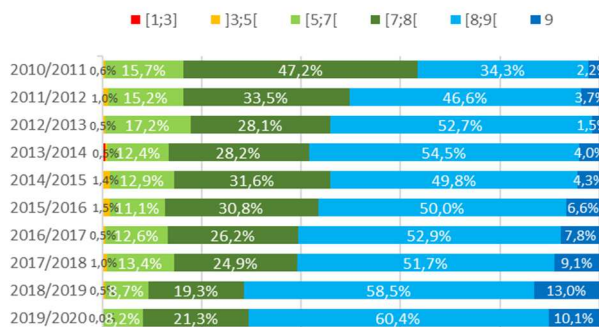


Fig. 12. Pedagogical performance of female teachers by the weighted RADIST classification.

We can see the most accentuated increase in this ‘improving towards excellency’ trend in male teachers with “excellent” results in 2016/2017 (compared to previous years). We can also notice an increase of 1% every year until 2019/2020, accompanied by a significant decrease in cases with “regular +” results and a less significant of “improving” cases.

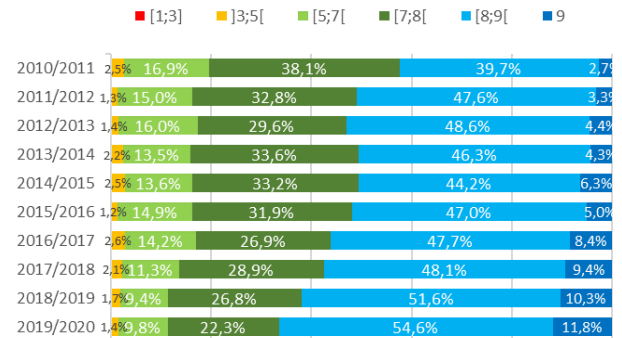


Fig. 13. Pedagogical performance of male teachers by the weighted RADIST classification.

When analyzing the performance of teachers by age group (Fig. 14 to Fig.18), it is also possible to perceive the same global trend by school year, however some evidences can be highlighted in each age group.

The pedagogical performance of teachers aged between 20 and 29 years is very good, in general, with all having a weighted RADIST classification above 7 and an unusual situation in 2013/2014 with 2.6% of cases with “inadequate” results (Fig. 14).

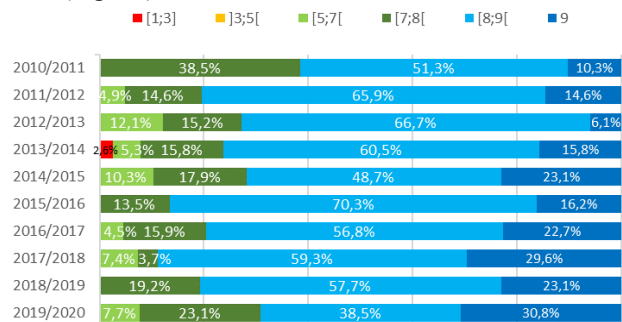


Fig.14. Pedagogical performance of teachers aged between 20 and 29 years, according to the weighted RADIST classification.

The pedagogical performance of teachers aged between 30 and 39 years, mostly shows results above “regular”, but there are more cases, compared to younger teachers, with performance “to improve” or even “inadequate” (Fig. 15).

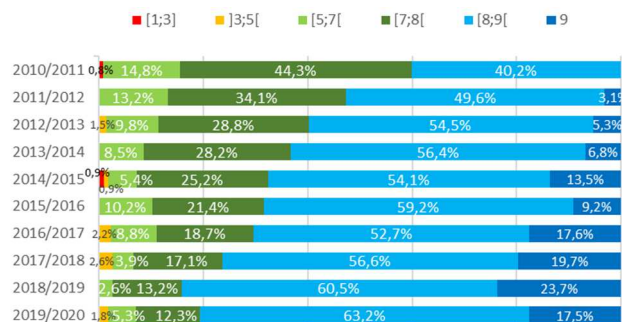


Fig. 15. Pedagogical performance of teachers aged between 30 and 39 years, according to the weighted RADIST classification.

The pedagogical performance of teachers aged between 40 and 49 years, also shows some cases of pedagogical performance “to improve” or “inadequate” over the years, despite the great majority having results above “regular” and with the same growing trend of “excellent” results, despite the reversal of this trend in 2019/2020 (Fig. 16).

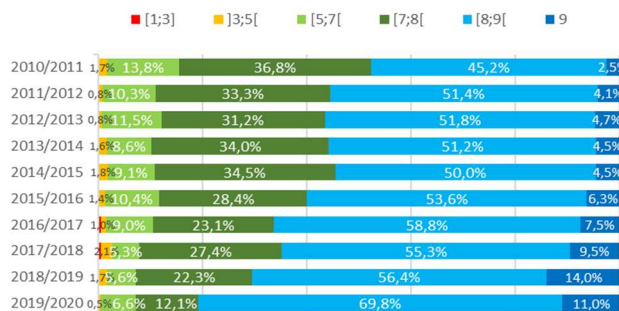


Fig. 16. Pedagogical performance of teachers aged between 40 and 49 years, according to the weighted RADIST classification.

In the case of teachers aged between 50 and 59 years, in contrast to younger teachers, a lower proportion of teachers with “excellent” results can be seen, but still with the same growing trend over the years. The cases with “improving” results, although more frequent, have been decreasing over the years - in 2010/2011 they were 4.4% and in 2019/2020 they were only 0.8%. (Fig. 17).

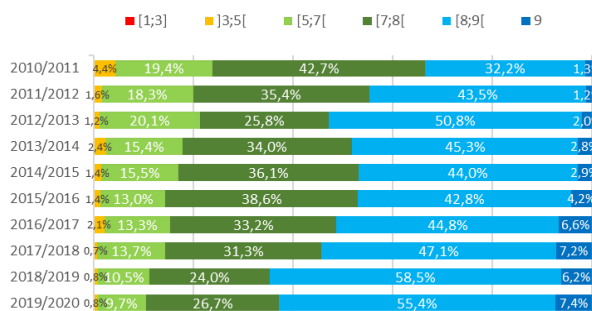


Fig. 17. Pedagogical performance of teachers aged between 50 and 59 years, according to the weighted RADIST classification.

The older teachers are those who reveal fewer cases with “excellent” results and who are more or less equally distributed between the “regular”, “regular +” and “very good” categories and, therefore, with more cases with “improving” results, which in 2014/2015 reached a maximum of 7.1% of cases compared to all other academic years and all other age groups (Fig. 18).

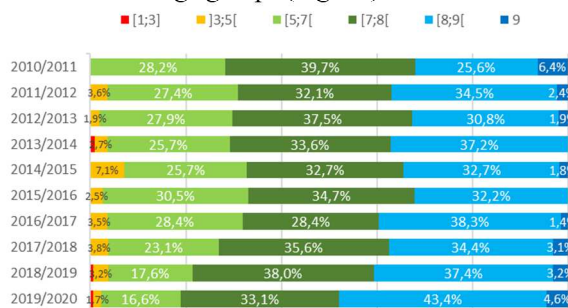


Fig. 18. Pedagogical performance of teachers aged 60 years, or over according to the weighted RADIST classification.

Regarding nationality, most teachers are Portuguese. Regarding the category, most of them have a “Professor”

category and, therefore, the teaching performance of teachers per academic year will only be presented graphically for the categories of “Assistant Professor” (40-52%), “Associate Professor” (18-23%) and “Full Professor” (9-10%).

Teachers that have category of Assistant Professor are the most representative in the universe of professors evaluated by the QUC and show a very good pedagogical performance in the majority. In the last year, the percentage of teachers with an “excellent” result decreased by about 3% compared to the previous year, although it wasn’t identified cases with performance “to improve” and only one exceptional “inadequate” case (0.3%), a value not registered in any other academic year of the analysis.

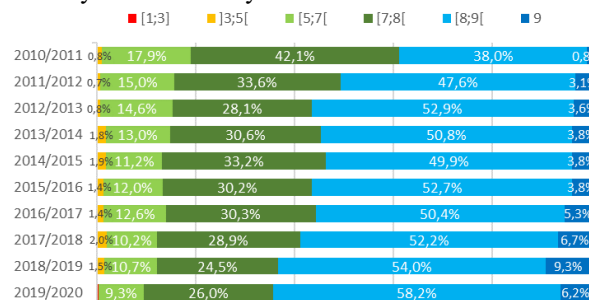


Fig. 19. Pedagogical performance of teachers with the category of Assistant Professor, according to the weighted RADIST classification.

Teachers in the category of Associate Professor have had, over the years under analysis, a pedagogical performance that is mostly regular to very good, and since 2016/2017, more than half have been showing very good to excellent results, which in the latter case registered a significant increase in 2019/2020 compared to the previous five years, reaching almost 9% of “excellent” cases (Fig. 19). In contrast, despite this improvement, there was an increase, even if residual, of cases in this category with “to improve” pedagogical performance between 2017/2018 and 2019/2020, being 2.2% in the latter academic year.

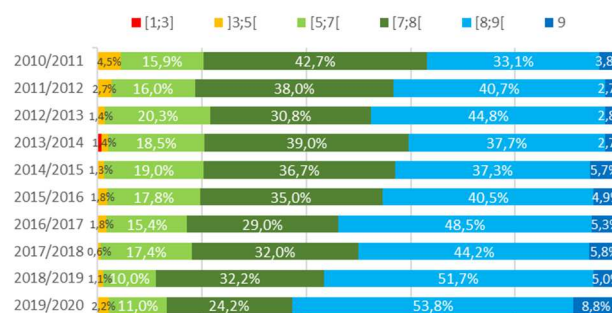


Fig. 20. Pedagogical performance of teachers with the category of Associate Professor, according to the weighted RADIST classification.

Although less representative in the teaching staff, Full Professors mostly show a regular to very good pedagogical performance. The cases with “excellent” results stand out in the last academic years, despite the decrease registered in 2018/2019. In contrast, the cases with results “to improve” reached its maximum in 2017/2018 (5.3%), but in the immediately following year was non-existent, emerging 1.3% of teachers in 2019/2020 (Fig. 21).

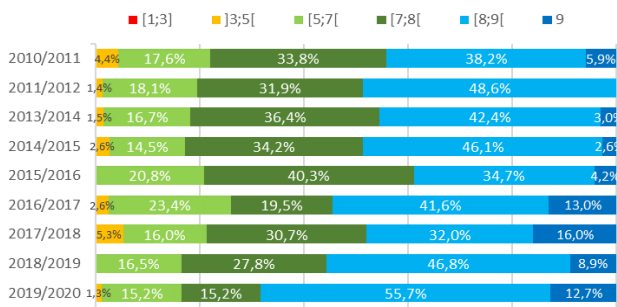


Fig. 21. Pedagogical performance of teachers with the category of Full Professor, according to the weighted RADIST classification.

Focusing the analyses on the distribution of the pedagogical performance of the teachers integrated in the groups of the programs of observation of classes, in the moments (years) of observation, it is possible to perceive the differences by academic years and in within different groups (Fig.22 and 23).

In Group A, teachers were observed between 2015/2016 and 2019/2020, and they were not necessarily observed in consecutive years, as can be seen by the universe of teachers belonging to this group by academic year (Table I).

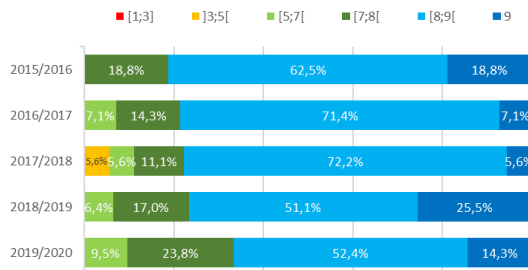


Fig. 22. Pedagogical performance of teachers observed in Group A, according to the weighted RADIST classification.

Most teachers in Group A have a very good to excellent pedagogical performance, with the rest having a regular performance, with a single exceptional case with performance “to improve” in 2017/2018 (Fig.22).

In Group B, teachers were observed between 2012/2013 and 2017/2018, and they were not necessarily observed in consecutive years, as can be seen in the universe of teachers belonging to this group by academic year (Table I).

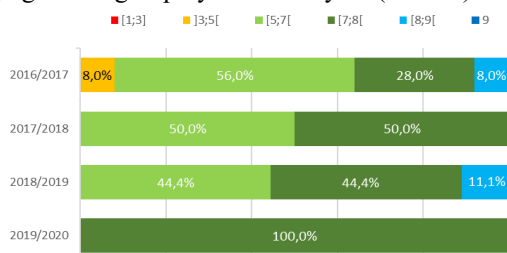


Fig. 23. Pedagogical performance of teachers observed in Group B, according to the weighted RADIST classification.

The teachers observed in Group B have a more average pedagogical performance, with a RADIST classification between 5 and 8. However, the “very good” cases in 2016/2017 (8%) and in 2018/2019 (11%) stand out. It can be said that there has been a slight improvement trend over the years, considering that in the last year only one teacher was

observed (Fig. 23). In Group C of classes’ observation, the teachers had a regular pedagogical performance, that is, with a weighted RADIST classification between 5 and 8.

## VI. ANALYSIS OF TEACHING PEDAGOGICAL PERFORMANCE AND IMPACT OF CLASS OBSERVATION

The main objective of this study, was motivated by a set of questions, such as: “Are there any differences in the RADIST classification between the different evaluation moments? Depending on whether or not you were enrolled in one of the Groups of Class Observations? Is there any effect of the “Class Observation” factor on the pedagogical performance of teachers? And, finally, if there is any effect, will the program / Does the group in which the teachers were observed have any impact on this effect? ”. In order to get some answers, several statistical tests were applied.

These and other questions can be explored in various ways, such as, for example, comparing the pedagogical performance of teachers evaluated based on the weighted RADIST classification in two, out of three, different moments; compare the pedagogical performance of the teachers evaluated between two different populations, based on the Observation/No Observation from classes, at two different times; to compare the pedagogical performance of the teachers evaluated between two different populations, based on the group of Class Observation in which the teachers were enrolled, in two different moments; all of these can be tested by using the Wilcoxon's Non-Parametric Tests [2].

Based on the RADIST classification obtained under the QUC process, the information was organized by each teacher evaluated in each academic year considered in the analysis referring:

- Whether or not the teacher was enrolled in Class Observation(s) in the respective academic year, (regardless of the semester(s) in which the observation occurred);
- The group where the teacher was enrolled in (A, B or C), if participated in the observations within the scope of any of the Class Observation programs (s);
- The weighted RADIST classification in three different moments:
  - Before that academic year of evaluation, if any;
  - In the academic year of evaluation, if any;
  - After the academic year of evaluation, if any.

We can see the distribution of the weighted RADIST classification at the different moments described, for the “Before”, “Present” and “After Moments” (Fig. 24, 25, 26). In the following figures it is possible to observe that in any of the moments, the pedagogical performance of the teachers approaches an asymmetric distribution on the left, with an average around the value of 7.8 in each of the moments. In spite of not approaching a normal distribution, since the samples have a dimension much larger than 30, the normality of the data can be assumed and, thus, proceed with the application of the mentioned statistical tests.

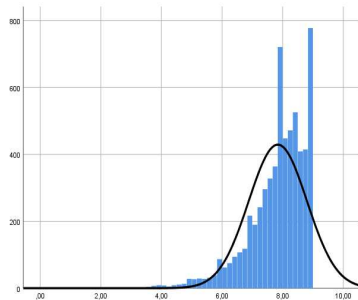


Fig. 24. Distribution of the number of RADIST classifications weighted in the “Before” moment, in the years under analysis.

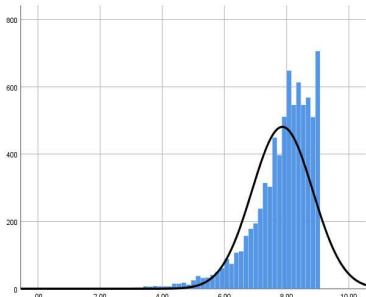


Fig. 25. Distribution of the number of RADIST classifications weighted in the “Present” moment, in the years under analysis.

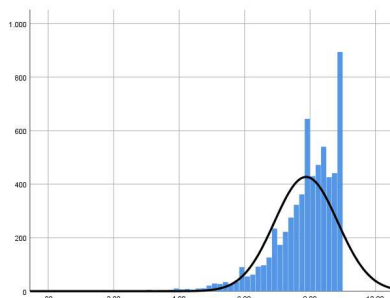


Fig. 26. Distribution of the number of RADIST classifications weighted in the “Present” moment, in the years under analysis.

It is intended to perceive not only the evolution of the pedagogical performance of teachers over time, in three different moments, through the measure of pedagogical performance (weighted RADIST classification) but also understand if this process has had any impact on their pedagogical performance.

Recalling the issues set out in the framework, a Wilcoxon non-parametric statistical test was performed to compare the pedagogical performance, based on the weighted RADIST classification, of the same teacher at two different times:

- “Before” and in the “Current” moment
- “Currently” and “After”
- “Before” and in the moment “After”

Based on the test results conducted, it is possible to conclude that there is statistical evidence that **there are differences between the teaching performance of teachers over time**. This fact has already been verified by the graphic representations of the measure over time, which show an increasing trend of teachers with better pedagogical performance (“very good” and “excellent”).

When carrying out the same analysis for two different populations: those who were enrolled in Class Observations and those who were not enrolled in Class Observations («Observed» or «Not Observed»), we can conclude, based on the test results the following:

- «Not Observed»: there is statistical evidence to affirm that there are differences between the pedagogical performance of teachers over time.
- «Observed»: there is statistical evidence to affirm that there are differences between the pedagogical performance of teachers in the “Before” and the “Current” moments; and in the “Before” and the “After” moments.
- On the other hand, there is no statistical evidence to affirm that there are differences between the pedagogical performance of teachers in the “Current” and the “After” moments.

Therefore, if there is a statistically significant difference between the pedagogical performance of the teachers who were enrolled in Class Observations in two of the tests. It is also interesting to notice at which moment the median of the pedagogical performance measure is higher.

Despite the evidence, it is clear that the difference identified between the two moments in both cases is minimal and, therefore, it is not possible to draw a strong conclusion about the effect of class observation on the performance of teachers, since the performance trend of teachers has followed a path of improvement over the years.

Looking for more answers, the same test was also carried out for each of the groups for class observation with a significant number of teachers, thus considering Groups A and B and excluding Group C. The test intends, within the same group of class’ observations, to analyze whether there are differences between the teachers’ pedagogical performance, clarifying if the differences identified in previous tests (that there is an effect of classroom observation), are somehow related to the class observation group.

Based on the results of the tests applied to the teachers observed in Group A, it is observed that there is statistical evidence to reject the equality of the median of the pedagogical performance measure of the teachers of Group A in the “Before” and in the “After” moments, which was 8.37 in the moment “Before” and 8.51 in the moment “After”. Despite this difference, and as previously seen in the characterization of this group of teachers, it is a group with mostly “very good” pedagogical performance that also follows the global trend of teachers in improving their pedagogical performance over the years.

Analyzing the results of the tests applied to the teachers observed in Group B, it is possible to state that there is statistical evidence to reject the equality of the median of the measure of the pedagogical performance of the teachers in the moment “Before” the observation (6.46) and in the “Current” moment (6.81). Despite the difference, and as previously seen in the characterization of this group of teachers, in addition to

being a group with mostly “regular” pedagogical performance, it also does not show an evolutionary pattern of performance, perhaps because fewer observations were made during the analysis period and, therefore, it is more difficult to corroborate this improvement conclusion between these moments.

In order to better understand the behavior of the pedagogical performance of the teachers observed, by any of the Class Observation groups (A, B or C), depending on the three different moments, three different situations were represented graphically (Fig. 27, 28, 29). It should be noted that in the vertical and horizontal lines, with green color, it is represented the minimum limits of the RADIST classification of the groups designated as “Regular” and “Regular +” (check II in this paper), which facilitates the identification of the highest concentration of cases particularly in the analysis of each graphic representation, above or below each of these limits.

The distribution of the weighted RADIST rating of teachers at the “Current” moment (year of Class Observation(s)) according to the weighted RADIST rating of teachers at the moment “Before” the observation occurred, reveals a certain linear relationship, especially in cases above 7. It seems that the better the performance before the observation of classes, the better teacher’s performance will be in the year in which the observation takes place. In cases below 7, this trend is not so evident since the cases are more dispersed and can either improve or not their performance in the year of observation (Fig.27).

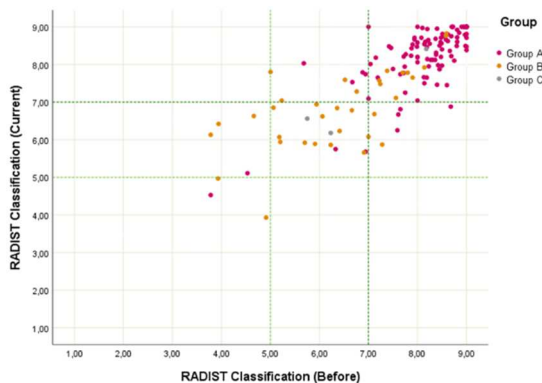


Fig. 27. Distribution of the RADIST Classification weighted in the “Current” moment according to the RADIST Classification weighted in the “Before” moment of the Teachers' Class Observation, by Group of Observation.

The distribution of the weighted RADIST rating of teachers at the “Current” moment (year of Class Observation(s)) compared to the weighted RADIST rating of teachers at the time “After” the observation has occurred, reveals something similar to the previous case, that is, a linear relationship between the two moments mainly in cases above 7. The remaining cases are more dispersed, with some that at the moment of observation have a RADIST classification between 5.5 and 7 and later (moment after observation) a better performance, between 7 and 8, approximately. Just as there are also some cases that performed better (RADIST > 7) at the moment of observation and subsequently slightly decreased their performance with values between 6 and 7,

approximately (Fig. 28). Despite being few cases, they continue to have their relevance.

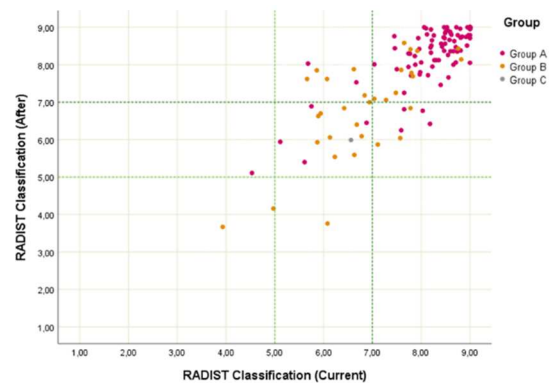


Fig. 28. Distribution of the RADIST Classification weighted in the “After” moment according to the RADIST Classification weighted in the “Current” moment of the Teachers' Class Observation, by Group of Observation.

The distribution of the weighted RADIST rating of teachers at the moment “After” the observation of classes according to the weighted RADIST rating of teachers at the time “Before” the observation occurred, also reveals the most direct linear relationship in cases over 7 at both times. And, as previously verified, teachers with a RADIST classification “Before” observing classes between 5 and 7 have a similar or superior performance (>6, mostly) at the “After” moment (Fig. 29).

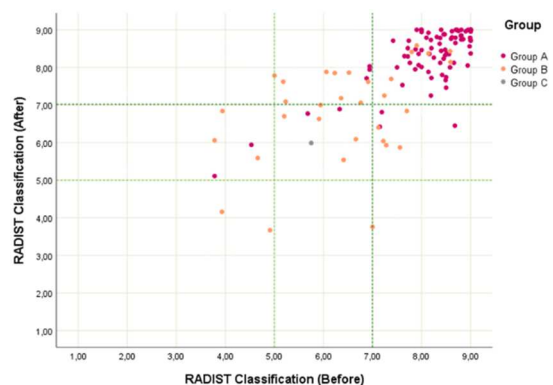


Fig. 29. Distribution of the RADIST Classification weighted in the “After” moment according to the RADIST Classification weighted in the “Before” moment of the Teachers' Class Observation, by Group of Observation.

## VII. CONCLUDING REMARKS

In this study, focusing on the universe of more than 700 teachers evaluated at Técnico annually by QUC, from 2010/2011 to 2019/2020, it appears that just over 1/4 are female; the representation by age group clearly identifies an aging of the teaching population, where 10 years ago there were 11.1% of teachers over 60 years old, in 2019/2020 they are already 21.9%, following the decrease of young teachers aged up to 39 years (until 8.7%). There are about 3% of teachers with foreign nationality since 2014/2015. It is clear that the majority of the teaching staff is composed by Assistant Professors (~25%) or Associates (~50%) every academic year, revealing a growing trend for the last mentioned at the expense of the decrease in Assistant Professors since 2012/2013.

From the same universe of teachers, 116 teachers were observed between 2015/2016 and 2019/2020 by the Group A classroom observation program, 37 by the Group B between 2016/2017 and 2019/2020, and only 3 by the Group C in the last two academic years. The tendency of the profile indicators of teachers is slightly different between groups and in comparison with *Técnicos*'s overall. In the group A, female teachers represent about 1/3, with the exception of two academic years, whereas in group B the proportion of female teachers is more variable over the years. Teachers observed in group A were mostly between 30 and 49 years old, in contrast to the teachers in group B who represent an older population. The most representative category of teachers observed in both groups A and B were the Assistant Professor, which meets the objectives of the Class Observation process at an earlier time in the teaching career, however in group B there are Full Professors with observed classes. Since group C is underrepresented, it is not possible to draw any conclusions.

From the analysis of the pedagogical performance of the universe of teachers considered in the study, the trend of improvement in performance over the years under analysis is evident, and this is maintained according to the gender of the teachers. However, observing the same by age group, this trend is clearest in those aged between 30 and 59 years, with the younger ones showing a very good to excellent performance, mostly, but with fluctuations over the years, and those with more years of teaching experience showing a mostly regular performance, with no evolutionary trend, at least until 2016/2017. The improvement trend is evident in the performance of Assistant and Associate Professors, however the Full Professors show a more fluctuating performance, but mostly very good to excellent in the last 4 years.

From the group of teachers observed in group A, it appears that they have a very good to excellent performance in the vast majority (above 70%) in all years, although there is no evidence of an improvement trend over the years. On the contrary, in group B teachers show mostly a regular performance without any evident tendency over the years.

As might be expected, the statistical tests applied corroborated some of the conclusions obtained from the graphic analysis of the above indicators, allowing for a finer and more temporal reading on the behavior of the teachers' performance during the period of analysis, making use of the effect of the classroom observation process.

It was then possible to conclude that there is statistical evidence to affirm that there are differences between the pedagogical performance of teachers over the years, namely with an emphasis on moments after the observation of classes has occurred, reinforcing the fact that the moment after may be in an immediately following year, or two more years ahead, depending on the occurrence of teacher evaluation within the scope of the QUC. More particularly, this difference reveals, based on the evolution of pedagogical performance, that the moment after the observation of classes, the pedagogical performance is better, i.e. it takes a value of the performance measure higher than in the previous one (year(s) before), even if this difference is by tenths.

This study proved to be important to keep monitoring the teaching performance of teachers for more years, as well as making more observations of classes to see if this factor influences or contributes to the improvement of the pedagogical skills of *Técnicos*'s teachers. In the future, it might be useful to see if observations lead to better results.

If considering more qualitative feedback, coming from the meetings of the teachers with NDA, class observations are seen as an appropriate means to promote the quality of teaching and the pedagogical development of teachers. The acceptability of this observation process at *Técnico* seems to relate to the constructive, formative and developmental role of the team who observes and the willingness of teachers to profit from the observation – in fact, “it becomes necessary to see observations as an opportunity for teachers to collaboratively engage in reflection on their professional performance and in the investigation and discussion of strategies that allow them to improve their practice” [1].

Within the scope of the observations made to the Professors and Teaching Researchers, there is a set of information that reflects, in a more striking way, the opinion about the positive impact that they consider the observations have in their teaching practice, following this link, only in portuguese: <http://shapingthefuture.tecnico.ulisboa.pt/shaping-the-future>.

The observation of classes' methodology described can also be useful for promoting the pedagogical development of teachers. Observations have been used to identify aspects of professional practice to be improved, to monitor progress and skills development, to provide contact and reflection on the potential and limitations of different approaches, to promote strategies and activities used in the classroom context or to develop and broaden horizons of the different dimensions of teachers' pedagogical knowledge (e.g. encouraging teachers to share their practices with their colleagues, both directly and indirectly through NDA team), building community and collective learning. We believe there is no exaggeration in stating that “excellent teachers are not born, they're made”, and they are made through incremental improvements based on self and other regulation, something that the observation of classes can be a useful tool to push forward.

#### ACKNOWLEDGMENT

The authors would like to acknowledge the Pedagogical and Scientific Council of *Técnico*, namely Prof. Raquel Aires de Barros and Prof. Luís O. Silva for their support and encouragement regarding the Observation of Classes.

#### REFERENCES

- [1] Reis, P., “*Observação de aulas e avaliação do desempenho docente*,” in *Cadernos do Conselho Científico para a Avaliação de Professores*, Vol. 2, June 2011. (in Portuguese)
- [2] Torres, A., David, F., Graça, M. “Quality assurance of teaching and learning: validity and usefulness of student ratings”. [EQAF 12<sup>th</sup> Annual Forum 2017]
- [3] Torres, A., Ribeiro, J.P., Graça, M., Barros, R.A. (2016). “Working in partnership with students: the importance of experience and engagement”. [EAIR 38th Annual Forum 2016]
- [4] Torres, A. “Engaging students in the teaching quality assurance: a driver of higher education excellence”. [EAIR 37th Annual Forum 2015]