Academics’ perceptions regarding performance evaluations and the consequences for their professional and personal activity

Javier Mula-Falcón and Katia Caballero
Department of Didactics and School Organisation, Faculty of Education, University of Granada, Granada, Spain

Abstract

Purpose – Improving and assuring the quality of higher education has become a key element of policy agendas worldwide. To this end, a complete accountability system has been developed through various evaluation procedures. Specifically, this study analyzes the perceptions of university teaching staff on the impact of performance appraisal systems on their professional activity, health and personal lives.

Design/methodology/approach – The study adopted a nonexperimental descriptive and causal-comparative design using a questionnaire that was completed by a sample of 2,183 Spanish teachers. The data obtained were analyzed using descriptive statistics and comparisons of differences.

Findings – The results show that, according to teachers, the evaluation criteria undermine the quality of their work by encouraging them to neglect teaching, increase scientific production and engage in unethical research practices. Their views also emphasize the social and health-related consequences of an increasingly competitive work climate, including increased stress levels. Finally, significant differences are observed regarding gender, professional category and academic discipline, with women, junior faculty and social sciences teachers expressing particularly strong views.

Originality/value – The originality of this study lies in the application of a method that contributes to the international debate through a national perspective (Spain) that has so far received little attention.

Keywords Quality, Accountability, Academics’ performance evaluations, Higher education, Academics perception, Bologna process, Consequences, Quantitative study

Paper type Research paper

1. Introduction

Improving and assuring the quality of higher education has become a central element of policy agendas worldwide, impacting all levels of the sector and becoming an essential, everyday element of academic life (Huisman et al., 2015; Browning et al., 2017). The aim of these changes was to achieve a quality higher education system that is globally competitive...
and capable of generating lasting economic development (Tomicic, 2019). Despite this, defining the term “quality” is complex as it is a polysemous, contextual concept with a significant ideological component (Tasopoulou and Tsiotras, 2017). Therefore, what we mean by quality will depend entirely on the actors who define it (Schindler et al., 2015). In this sense, Harvey (2006) asserts that quality assurance in higher education encompasses both enhancement (development) and control (accountability) and that the emphasis or balance between the two will depend directly on the governmental vision and the degree of autonomy and trust (Stensaker and Maassen, 2015).

Traditionally, quality assurance in higher education has focused on improvement, which has often taken the form of evaluation processes created by the institutions, making use of self-reports or peer review (Hazelkorn, 2018). However, changes in higher education (e.g. massification, specialization, globalization and new teaching challenges) have made these forms of evaluation scarce (Pattaro et al., 2022). As a result, recent decades have witnessed reforms aimed at improving the quality of the sector. This shift has necessitated the introduction of external evaluation processes (in most cases controlled by independent agencies and materialized in accreditation systems) focusing on various elements including curricula, mobility schemes and faculty (Stensaker, 2018). Moreover, these evaluation processes have been linked to certain incentives (e.g. funding, prestige, tenure or career advancement) (Dougherty and Natow, 2019). This system has developed into an accountability process with a clear objective of quality control, autonomy and performance (Macheridis and Paulsson, 2021; Huisman, 2018). Consequently, there has been a shift from a quality discourse focused on improvement to one focused on fostering control, competitiveness and comparability (Stensaker, 2018).

Despite this, current discourses uphold the belief that these evaluation processes are true instruments of transparency, democratic validity and quality improvement (Huisman, 2018; Gunn, 2018). However, the lack of evidence supporting the notion that these evaluation processes produce improvements (Pattaro et al., 2022; Macheridis and Poulson, 2021) leads us to consider them as a simple method of control, reward or punishment (Huisman, 2018; Gunn, 2018) that directly affects the staff involved (university faculty) and, therefore, their institutions (Dugas et al., 2020; Jiménez, 2019; Ursin et al., 2020; Huang et al., 2016). These teachers are required to undergo continuous evaluation processes that determine their academic trajectory (Ma and Ladisch, 2019; San Fabián, 2020; Laiho et al., 2020; Harvey, 2020). All this has consequences on work performance (prioritization of research over teaching, development of unethical practices, promotion of scientific production), on social and family relations (sacrifices of family and free time, problems with parenthood, etc.) and on the health of university teaching staff (stress, anxiety, etc.) (Bolivar and Mula-Falcón, 2022; Bozzon et al., 2017; Acker and Webber, 2016; El-Far et al., 2020; Lipton, 2020; Hortal and Tang, 2023; Harris et al., 2019). Consequently, Saura and Bolivar (2019) point out the enormous “manipulative” power of these evaluations since the continuous quest to meet these criteria has a profound impact on teachers’ ways of doing. In other words, the struggle to pass these assessments leads teachers to do whatever the assessment criteria demand, meaning that their academic output will largely depend on such criteria.

Therefore, in this study, we sought to ask university teaching staff about their views regarding the impact of these performance evaluations on the development of their professional activity, health and personal life. In other words, we wanted to determine if these accountability systems truly promote improvement and development or, on the contrary, whether they run counter to the objectives, values and functioning of the university, negatively impacting the quality of work and promoting job insecurity, competitiveness and individualization (Djerasimovic and Villani, 2019; Tülübass and Göktürk, 2020). Furthermore, we aimed to establish whether these perceptions differ according to three
fundamental variables: gender, professional category and academic discipline. Therefore, the
main objectives of this study are as follows:

(1) To determine university teachers’ perceptions of the assessment criteria used to
evaluate their professional activity.

(2) To describe the main influences/consequences generated by the evaluation criteria on
their professional and personal activity.

(3) To determine the differences in the perceptions of the evaluation criteria according to
certain demographic variables (gender, professional category and academic
discipline).

(4) To determine the differences in the perceived impact of the criteria on professional
and personal activity according to various demographic variables (gender,
professional category and academic discipline).

A quantitative study was conducted on a sample of Spanish teachers to answer these
objectives. We focus on the Spanish context for several reasons. First, few studies have been
conducted on Spanish teachers in this field of research (Mula-Falcón and Caballero, 2022).
Second, the Spanish higher education system has a long tradition (more than 20 years) of
developing and applying accountability processes due to the influence of the Bologna Plan.
Third, because Spain is in a period of transition in its university policy in which scientific
evidence is needed to contribute to the social and political debate. Furthermore, the
quantitative approach was selected because (1) there is a scarcity of studies using this
approach (Mula-Falcón and Caballero, 2022) and (2) it allows obtaining more comprehensive
and objective data that contribute to making inferences and comparisons between contexts.

To this end, in this article, we will first describe the evaluation procedures used for
Spanish university teaching staff. The methodology will then be explained, and the results
will be presented. Finally, the results will be discussed, and a series of conclusions and future
lines of research will be proposed. By meeting the objectives described above, we aim to (1)
delve deeper into the state of this issue in the Spanish context; (2) contribute to the
international debate through a national perspective that has so far received little attention;
and (3) contribute to two of the sustainable development goals, namely, SDG 4 (quality
education) and SDG 8 (decent work and economic growth).

2. The quest for quality in the Spanish context: accreditations
The Bologna Plan – or Bologna Process – was designed to achieve a complex transformation
of the European university system that began in 1999. This process aimed to promote the
homogenization of European Higher Education by developing a common framework
(European Higher Education Area), which would allow educational harmonization,
comparability and free movement of students or employability, among other aspects
(Brogger, 2019).

One of the main pillars of the Bologna Plan was to guarantee the quality of the European
university system in all areas, including teaching, research, mobility, teaching staff
and institutions. To this end, and in line with international policy, a series of agreements and
treaties were developed to create evaluation designs, criteria, methodologies, guidelines
and certification mechanisms to ensure the quality of the European higher education system.
Furthermore, this process led to the development of national accreditation agencies following
the guidelines proposed by the European Association for Quality Assurance in Higher
Education, a European network aimed at promoting the enhancement and maintenance of
quality in higher education.
From the outset, Spain has been immersed in this process of European convergence, which began to materialize in 2001 with the Organic Law on Universities and in 2002 with the creation of the National Agency for Quality Assessment and Evaluation (ANECA in its Spanish acronym). The latter is the Spanish national accreditation agency responsible for evaluating the university system. The roles of this system include the evaluation of study plans (VERIFICA program), the monitoring and implementation of official degrees (MONITOR and ACREDITA programs) and the implementation of doctoral programs (MENTION program). In addition, ANECA manages the evaluation of university teaching staff through the PEP, ACADEMIA and CNEAI programs.

PEP and ACADEMIA are accreditation programs that evaluate university teaching staff for all positions in the Spanish university system. The PEP program evaluates assistant and contracted lecturers, while the ACADEMIA program evaluates tenured and tenure-track lecturers. The outcome of these evaluations will determine access to the university system (in the case of the assistant and contract categories) or career progression through the rest of the levels (in the case of tenured and tenure-track professors).

These evaluations focus on teaching, research, management and transferable functions; professional experience outside the university environment; and training. However, the percentage weightings given to each of these components vary depending on the program. Thus, for the PEP, research work accounts for 60% of the total evaluation, in which priority is given to scientific production in impact journals. On the other hand, in the ACADEMIA program, although teaching and research work are given preference over the other components, both carry equal weighting. However, it is not necessary to obtain an excellent score on both components to pass the evaluation.

Finally, the CNEAI is a program that evaluates the research activity of university teaching staff for six-year periods and in which priority is given to scientific production in impact journals. A favorable evaluation translates to salary incentives and preferential merit for passing the ACADEMIA program.

3. Method
A quantitative methodology with a nonexperimental design was employed to respond to these objectives. This method was selected for two reasons: (1) to provide a first general descriptive overview of this field of study in the Spanish context and (2) to offer a quantitative overview that allows its extension to other international contexts (Mula-Falcón y Caballero, 2022). Specifically, a descriptive design was used for the first and second objectives, while a causal-comparative design was used for the last two. The aim was to obtain an overall view of the object of study and establish possible differences according to gender, professional category and academic discipline.

3.1 Participants
For the present study, all university teachers from nine Andalusian public universities (Spain) were invited to participate by e-mail. Initially, 2,588 teachers responded to the questionnaire, of which 405 were excluded because the responses were incomplete or were identified as erroneous by the system. The final sample consisted of 2,183 teachers, representing 12.4% of the population, with a sampling error of 1.96 for a confidence interval of 95%.

Table 1 summarizes the number of teachers surveyed from each of the universities. Concerning the distribution of the sample according to university, no statistically significant differences were found ($p = 0.72$) when applying the Mann–Whitney $U$ test.
Of this sample, 53.7% identified as male, 45% as female and 1.2% identified as other or preferred not to answer. Concerning professional category, 29.4% were first-level professors (predoctoral, postdoctoral and temporary substitute professors), 19.4% were second-level professors (assistant and contract professors), 51% were third-level professors (tenured and tenure-track professors) and 0.1% of the sample did not indicate their professional category. Finally, regarding academic discipline, 17.9% worked in arts and humanities, 20% sciences, 15.5% health sciences, 30.6% social and legal sciences, 15.9% engineering and architecture and only 0.1% did not indicate the discipline to which they belonged.

3.2 Instruments
The instrument employed was an unpublished questionnaire developed ad hoc for the national research project R + D + I funded by the Ministry of Science and Innovation of the Government of Spain, in which the present study is framed. For the elaboration of this questionnaire, an in-depth review of the literature was carried out (ANONIMOUS). This review revealed the existence of other questionnaires focusing on teachers’ general perceptions of their profession, on the training received and on organizational changes in the profession (Ylijoki et al., 2011; Teichler et al., 2013; Santos et al., 2015; Zenatta et al., 2017). However, no studies have been found with questionnaires on teachers’ perceptions about the evaluation of their professional performance. Therefore, the need arose to create an instrument to find out how teachers perceive and feel the evaluation systems of their professional performance (in Spain evaluations carried out by PEP, ACADEMIA and CNEAI programs) and how they affect their professional activity and their personal life. The questionnaire created was entitled Perceptions and Satisfaction of university teaching staff on the Development and Evaluation of their professional activity (PSPU in its Spanish acronym) and was subjected to both validity (expert judgment) and reliability (exploratory factor analysis) analyses.

In relation to the validity analysis, the Kendall’s concordance coefficient allowed us to determine that the ten experts (both specialists in the subject matter under study and members of evaluation agencies with in-depth knowledge of the subject matter) rated the different dimensions of the instrument positively. On the Relevance scale, Kendall’s W value was 0.436, and on the Clarity scale: 0.284, both values allowed us to reject Ho for an X^2_92 = 196.34 (Relevance) and an X^2_92 = 125.30 (Clarity), with a degree of significance of <0.001.

As a result, the final questionnaire consisted of 42 items distributed across 4 factors that aim to measure the perception and satisfaction of university teaching staff concerning the criteria for evaluating their professional activity. The reliability of the total number of items,
measured by Cronbach’s alpha, is 0.937. On the other hand, exploratory factor analysis showed that the set of factors explained 45.68% of the variance.

Only the responses to two dimensions were considered for this study. Dimension 1 analyzes teachers’ perceptions of the assessment criteria, while Dimension 2 analyzes the influence of the criteria on teachers’ professional and personal activity. These dimensions are composed of 9 and 11 items, respectively, which are answered using a Lickert-type scale with seven response options (1 = completely disagree/7 = strongly agree).

The reliability indices obtained for each dimension using Cronbach’s alpha were 0.815 and 0.765. According to Tavakol and Dennick (2011) and Serbetar and Sedlar (2016), values above 0.70 are considered acceptable and confirm the internal consistency of this questionnaire.

3.3 Data analysis
The results were analyzed using various statistical methods. First, a descriptive analysis was conducted to address the first two objectives by calculating means, standard deviations and frequencies. For the last two objectives, pairwise contrasts were used to determine differences. For this purpose, nonparametric tests were used after rejecting the null hypothesis ($p < 0.001$) using a Kolmogorov–Smirnov analysis, i.e., there was no normal distribution. The nonparametric tests used were the Mann–Whitney $U$ test for dichotomous variables (gender) and the Kruskal–Wallis test for nondichotomous variables (professional category and academic discipline). SPSS software was used for data analysis and calculations.

3.4 Ethical considerations
This study was approved by the Ethics Committee of the researchers’ university. Participation was, therefore, voluntary and based on the principles of informed consent in which the information collected is restricted to use for research purposes only and in which anonymity and confidentiality are assured.

4. Results
4.1 Descriptive analysis
4.1.1 Dimension 1: teachers’ perception of assessment criteria. Table 2 shows the scores representing the teacher’s perceptions of the assessment criteria. Within this, Items 16 (The assessment criteria for accreditation or six-year periods determine the way I direct my academic activity) and 25 (The priority given to research merits in the assessment processes harms my teaching activity) show the highest mean scores (4.38 and 4.30), with a median of 5. Item 30 (Research evaluation criteria encourage my production volume) is also notable with a mean of 4.06 and a median of 4. These items indicate that, according to the teachers’ perceptions, the criteria guide professional activity in a way that negatively impacts their teaching to prioritize academic outputs.

On the other hand, the rest of the items obtained mean scores below 4. Of the items with lower values, Item 18 (The evaluation criteria have a positive impact on the improvement of my academic activity) stands out with a mean of 2.86 and a median of 3. In addition, the scores of Item 19 (The evaluation criteria for accreditation or six-year periods favor cooperation between colleagues) stand out, showing the lowest mean (2.57) and median of the whole dimension (1). These low scores indicate how teachers believe that the criteria negatively affect peer cooperation.

4.1.2 Dimension 2: influence of the criteria on university teaching staff’s professional and personal activity. Dimension 2 concerns the influence of the assessment criteria on professional and personal activity (Table 3). The summary item shows a mean of 5.04,
16. The evaluation criteria for accreditation and six-year periods determine the way I direct my academic activity
17. The evaluation criteria for accreditation and six-year periods adequately assess my professional quality
18. The evaluation criteria have a positive impact on the improvement of my academic activity
19. The assessment criteria for accreditation and six-year periods favor cooperation between colleagues
24. The teaching evaluation criteria favor my involvement in improving my students' learning
25. The priority given to research merits in the evaluation process has a negative impact on my teaching activity
29. The research evaluation criteria have a positive impact on the quality of my research activity
30. The research evaluation criteria incentivize my academic output
31. The research evaluation criteria contribute to the rigor of my research processes

\*Source(s): Authors' own creation\*

<table>
<thead>
<tr>
<th></th>
<th>1 R (%)</th>
<th>2 R (%)</th>
<th>3 R (%)</th>
<th>4 R (%)</th>
<th>5 R (%)</th>
<th>6 R (%)</th>
<th>7 R (%)</th>
<th>X (SD)</th>
<th>Mdn</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>377 (17.3)</td>
<td>200 (9.2)</td>
<td>157 (7.2)</td>
<td>267 (12.2)</td>
<td>318 (14.6)</td>
<td>391 (17.9)</td>
<td>473 (21.7)</td>
<td>4.38 (2.159)</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>583 (26.7)</td>
<td>371 (17.0)</td>
<td>367 (16.8)</td>
<td>346 (15.8)</td>
<td>268 (12.3)</td>
<td>167 (7.7)</td>
<td>81 (3.7)</td>
<td>3.08 (1.779)</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>687 (31.5)</td>
<td>393 (18.0)</td>
<td>348 (15.9)</td>
<td>322 (14.8)</td>
<td>230 (10.5)</td>
<td>133 (6.1)</td>
<td>70 (3.2)</td>
<td>2.86 (1.747)</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>859 (39.3)</td>
<td>407 (18.6)</td>
<td>291 (13.3)</td>
<td>282 (12.9)</td>
<td>164 (7.5)</td>
<td>134 (6.1)</td>
<td>46 (2.1)</td>
<td>2.57 (1.703)</td>
<td>2</td>
</tr>
<tr>
<td>24</td>
<td>563 (25.8)</td>
<td>378 (17.3)</td>
<td>283 (13.0)</td>
<td>342 (15.7)</td>
<td>252 (11.5)</td>
<td>253 (11.6)</td>
<td>112 (5.1)</td>
<td>3.25 (1.904)</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>417 (19.1)</td>
<td>225 (10.3)</td>
<td>150 (6.9)</td>
<td>260 (11.9)</td>
<td>253 (11.6)</td>
<td>378 (17.3)</td>
<td>500 (22.9)</td>
<td>4.30 (2.232)</td>
<td>5</td>
</tr>
<tr>
<td>29</td>
<td>495 (22.7)</td>
<td>309 (14.2)</td>
<td>289 (13.2)</td>
<td>374 (17.1)</td>
<td>320 (14.7)</td>
<td>274 (12.6)</td>
<td>122 (5.6)</td>
<td>3.47 (1.900)</td>
<td>3</td>
</tr>
<tr>
<td>30</td>
<td>409 (18.7)</td>
<td>231 (10.6)</td>
<td>203 (9.3)</td>
<td>302 (13.8)</td>
<td>340 (15.6)</td>
<td>419 (19.2)</td>
<td>279 (12.8)</td>
<td>4.06 (2.068)</td>
<td>4</td>
</tr>
<tr>
<td>31</td>
<td>553 (25.3)</td>
<td>319 (14.6)</td>
<td>284 (13.0)</td>
<td>296 (13.6)</td>
<td>296 (13.6)</td>
<td>271 (12.4)</td>
<td>164 (7.5)</td>
<td>3.43 (1.994)</td>
<td>3</td>
</tr>
</tbody>
</table>

**Table 2.** Scores per item on Dimension 1

Academics’ perceptions
<table>
<thead>
<tr>
<th></th>
<th>1 R (%)</th>
<th>2 R (%)</th>
<th>3 R (%)</th>
<th>4 R (%)</th>
<th>5 R (%)</th>
<th>6 R (%)</th>
<th>7 R (%)</th>
<th>X (SD)</th>
<th>Mdn</th>
</tr>
</thead>
<tbody>
<tr>
<td>21. I spend too much time on the bureaucratic-administrative aspects of my academic activity</td>
<td>115 (5.3)</td>
<td>111 (5.1)</td>
<td>137 (6.3)</td>
<td>208 (9.5)</td>
<td>237 (10.9)</td>
<td>442 (20.2)</td>
<td>933 (42.7)</td>
<td>5.47 (1.827)</td>
<td>6</td>
</tr>
<tr>
<td>26. My participation in teaching innovation projects, congresses, and courses on teaching is due to the need to gather merit</td>
<td>447 (20.5)</td>
<td>256 (11.7)</td>
<td>186 (8.5)</td>
<td>291 (13.3)</td>
<td>251 (11.5)</td>
<td>335 (15.3)</td>
<td>417 (19.1)</td>
<td>4.06 (2.201)</td>
<td>4</td>
</tr>
<tr>
<td>27. My participation in teaching innovation projects, congresses, and courses is motivated by an interest in improving my teaching</td>
<td>267 (12.2)</td>
<td>157 (7.2)</td>
<td>169 (7.7)</td>
<td>318 (14.6)</td>
<td>328 (15.0)</td>
<td>469 (21.5)</td>
<td>475 (21.8)</td>
<td>4.64 (2.005)</td>
<td>5</td>
</tr>
<tr>
<td>32. I pay attention to academic-professional platforms and networks (Google Scholar, ResearchGate, Academia.edu...) so that my publications have visibility and are cited</td>
<td>377 (17.3)</td>
<td>292 (13.4)</td>
<td>259 (11.9)</td>
<td>344 (15.8)</td>
<td>306 (14.0)</td>
<td>298 (13.7)</td>
<td>307 (14.1)</td>
<td>3.93 (2.039)</td>
<td>4</td>
</tr>
<tr>
<td>35. My participation in research projects, conferences, and courses related to research is due to my interest in developing my research activity</td>
<td>128 (5.9)</td>
<td>90 (4.1)</td>
<td>85 (3.9)</td>
<td>218 (10.0)</td>
<td>339 (15.5)</td>
<td>597 (27.3)</td>
<td>726 (33.3)</td>
<td>5.40 (1.734)</td>
<td>6</td>
</tr>
<tr>
<td>36. Working to meet the evaluation criteria has a negative impact on my health (e.g. stress, anxiety)</td>
<td>204 (9.3)</td>
<td>154 (7.1)</td>
<td>115 (5.3)</td>
<td>206 (9.4)</td>
<td>297 (13.6)</td>
<td>421 (19.3)</td>
<td>786 (36.0)</td>
<td>5.13 (2.013)</td>
<td>6</td>
</tr>
<tr>
<td>37. I sacrifice aspects of my personal life (family, leisure, free time) to meet the demands of my academic activity</td>
<td>87 (4.0)</td>
<td>106 (4.9)</td>
<td>99 (4.5)</td>
<td>151 (6.9)</td>
<td>281 (12.9)</td>
<td>563 (25.8)</td>
<td>896 (41.0)</td>
<td>5.61 (1.688)</td>
<td>6</td>
</tr>
<tr>
<td>38. I feel guilty when I put aside my academic activity to devote time to myself or my family</td>
<td>324 (14.8)</td>
<td>195 (8.9)</td>
<td>135 (6.2)</td>
<td>219 (10.0)</td>
<td>307 (14.1)</td>
<td>438 (20.1)</td>
<td>565 (25.9)</td>
<td>4.63 (2.152)</td>
<td>5</td>
</tr>
<tr>
<td>39. Competitiveness in scientific production causes me personal discomfort and frustration</td>
<td>229 (10.5)</td>
<td>151 (6.9)</td>
<td>129 (5.9)</td>
<td>238 (10.9)</td>
<td>273 (12.5)</td>
<td>406 (18.6)</td>
<td>757 (34.7)</td>
<td>5.03 (2.054)</td>
<td>6</td>
</tr>
<tr>
<td>40. Dedication to family slows down or hinders my career advancement</td>
<td>361 (16.5)</td>
<td>271 (12.4)</td>
<td>186 (8.5)</td>
<td>298 (13.7)</td>
<td>272 (12.5)</td>
<td>332 (15.2)</td>
<td>463 (21.2)</td>
<td>4.24 (2.158)</td>
<td>4</td>
</tr>
<tr>
<td>41. The evaluation criteria for access and promotion at university make me consider changing job or searching for other professional opportunities</td>
<td>944 (43.2)</td>
<td>271 (12.4)</td>
<td>138 (6.3)</td>
<td>181 (8.3)</td>
<td>162 (7.4)</td>
<td>163 (7.5)</td>
<td>324 (14.8)</td>
<td>3.06 (2.298)</td>
<td>2</td>
</tr>
</tbody>
</table>

**Source(s):** Authors’ own creation
i.e., a medium–high level concerning the degree of agreement with the impact of the criteria on professional and personal activity. Moreover, it is worth noting that all items in this dimension show an average score above 3 (“agree”). Therefore, more than 70% of the teachers surveyed show a high degree of agreement regarding the influence of the assessment criteria on professional and personal activity. To determine the type of influence, we highlight the items with the highest and lowest mean.

Concerning the items with the highest scores in this dimension, two groups emerged. One group consisted of Items 21 (I spend too much time on the bureaucratic-administrative aspects inherent to my academic activity) and 35 (My participation in research projects, congresses, and courses related to research is due to my interest in developing my research activity) with a mean of 5.47 and 5.40. These items indicate the most significant work-related consequences of the evaluation criteria. Thus, Item 21 refers to the bureaucratic overload of the profession, which is accentuated in the evaluation processes. And Item 35 indicates how, among the items related to participation in different projects, congresses and courses (Items 26, 27 and 35), those related to research are the most highly valued by the participants due to their consequences in the development and improvement of their research activity. These results may be due to the high weight given to research in the evaluation processes, the performance in which determines the likelihood of obtaining a high score.

On the other hand, a second group emerged with high scores on items associated with social and health consequences. In this group, we highlight Items 36 (Working to meet the evaluation criteria affects my health negatively: (stress, anxiety, . . .)), 39 (Competitiveness in terms of scientific production causes me personal discomfort or frustration), and 37 (I sacrifice aspects of my personal life (family, leisure, free time) – to meet the demands of my academic activity) with a median of 5 and averages of 5.13, 5.03 and 5.61, respectively. However, of these three items, Item 37 is notable. On this item, more than 40% of the sample indicates a high degree of agreement with the social and family sacrifices necessary to meet the demands of the current university system.

Finally, concerning the items with the lowest scores, we can highlight Items 32 (I pay attention to academic-professional platforms and networks (Google Scholar, ResearchGate, Academia.edu . . .) so that (my publications have visibility and are cited) and 41 (The evaluation criteria for access and promotion at university make me consider changing job or searching for other professional opportunities) with an average of 3.93 and 3.06. Item 41 is particularly noteworthy, with more than 40% of the sample indicating a low agreement concerning a change of profession. In other words, according to our sample, the evaluation criteria do not improve professional networks and platforms for increasing the dissemination of scientific work or the search for other career options outside academia.

### 4.2 Significant differences

Table 4 shows the degree of statistical significance found in the dimensions analyzed for the gender, professional category and academic discipline variables.

<table>
<thead>
<tr>
<th></th>
<th>Sex</th>
<th>Professional category</th>
<th>Scientific discipline</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>0.575</td>
<td>0.05*</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>D2</td>
<td>0.00*</td>
<td>0.000*</td>
<td>&lt;0.001*</td>
</tr>
</tbody>
</table>

**Note(s):** The significance level is 0.05  
*Reject the null hypothesis  
**Source(s):** Author’s own creation
According to the results, Dimension 1 (Perception of the assessment criteria) scores differ significantly according to professional category and academic discipline, as shown by values below 0.05. Concerning Dimension 2, the scores differ according to gender, professional category and academic discipline.

To further explore these significant differences, pairwise comparisons were conducted for each variable separately, and the results of which are presented below.

4.2.1 Gender. Concerning gender, significant differences were only found for Dimension 2, where women showed a higher mean rank higher than men (1218.73 vs 968.8). This aspect can be seen in Figure 1.

4.2.2 Professional category. Table 5 shows the pairwise comparisons for the professional category variable. Starting with Dimension 1, the post hoc tests revealed differences between advanced, beginner and intermediate categories. However, as can be seen in Figure 2, the intermediate categories showed the highest scores, that is, this group of academics shows highest values in the perception of the evaluation criteria than the advanced categories.

Table 5. Pairwise comparisons for the professional category variable

<table>
<thead>
<tr>
<th>Sample/sample</th>
<th>D1</th>
<th>D2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginner/intermediate</td>
<td>1,000</td>
<td>0.030*</td>
</tr>
<tr>
<td>Advanced/intermediate</td>
<td>0.012*</td>
<td>0.000*</td>
</tr>
<tr>
<td>Advanced/beginner</td>
<td>0.052</td>
<td>0.000*</td>
</tr>
</tbody>
</table>

Note(s): The significance level is 0.05

Beginner: predoctoral, postdoctoral, and replacement contracts; Intermediate: contract categories (assistant and contract doctor); Advanced: permanent categories (tenured and tenure-track)

Source(s): Author’s own creation
Concerning Dimension 2, significant group differences were found. However, these differences always favor the intermediate category, i.e., this group yielded the highest scores concerning the consequences generated by the criteria (Figure 3).

4.2.3 Academic discipline. The post hoc tests show differences in both dimensions for the pairs sciences–social and legal sciences and arts and humanities–social and legal sciences, and in Dimension 1 for the pair engineering and architecture–social and legal sciences (Table 6).

The pairwise comparison in dimension one shows how the differences always favor the discipline of social sciences (Figure 4). The same is true for Dimension 2 (Figure 5), i.e., the
differences in all cases always favor the social sciences. These results indicate that social sciences staff show a higher degree of agreement with the consequences generated by the assessment criteria.

5. Discussion
This study aimed to determine how Andalusian university teaching staff perceive the criteria used to evaluate their professional performance and the consequences of this accountability system. A further objective was to explore whether these perceptions differ according to gender, professional category and academic discipline. The results of our study have yielded several conclusions.

Academics at Andalusian public universities believe that the criteria for evaluating their professional performance impact their work performance, especially in teaching and research. Concerning research, our participants indicate that the evaluation criteria generate pressure to increase scientific production, understood as an increase in articles published in impact journals (SJR or JCR). According to many authors, this is a natural consequence of the

Table 6. Pairwise comparisons for academic discipline

<table>
<thead>
<tr>
<th></th>
<th>Sample-sample</th>
<th>D1</th>
<th>D2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sciences–arts and humanities</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Science–engineering and architecture</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Sciences–health sciences</td>
<td></td>
<td>0.708</td>
<td>0.075</td>
</tr>
<tr>
<td>Sciences–social and legal sciences</td>
<td></td>
<td>0.003*</td>
<td>0.001*</td>
</tr>
<tr>
<td>Arts and humanities–engineering and architecture</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Arts and humanities–health sciences</td>
<td></td>
<td>0.901</td>
<td>0.260</td>
</tr>
<tr>
<td>Arts and humanities–social and legal sciences</td>
<td></td>
<td>0.005*</td>
<td>0.011*</td>
</tr>
<tr>
<td>Engineering and architecture–health sciences</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Engineering and architecture–social and legal sciences</td>
<td></td>
<td>0.026*</td>
<td>0.164</td>
</tr>
<tr>
<td>Health sciences–social and legal sciences</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note(s): The significance level is 0.05
Significance values have been adjusted using the Bonferroni correction for several tests
Source(s): Author’s own creation

Figure 4. Pairwise comparisons for academic discipline (Dimension 1)

Source(s): Author’s own creation
current evaluation system in which scientific output has become a prerequisite for access, advancement, tenure and salary or professional benefits (Ma and Ladisch, 2019; San Fabián, 2020; Laiho et al., 2020). Moreover, faculty do not perceive that the criteria improve the quality or the rigor of their research, only the volume of their output. In this line, Harvey (2020) refers to the emergence of dubious ethical strategies as a consequence of the increased demands for scientific production, such as the use of “predatory” journals and publishers, “pay-to-publish,” exchange of authorship, duplication of studies or superfluous writing, among many others.

Concerning teaching, our participants have expressed the idea that the current criteria do not contribute to the development of their teaching quality or their involvement in improving their students’ learning. On the contrary, as Bolívar and Mula-Falcón (2022) state, the current university system contributes to relegating teaching to second place (or even dismissing it), considering it an obstacle that hinders those activities necessary to access more stable positions, i.e., the production of articles. Therefore, the overemphasis on research activity in the evaluation criteria generates an imbalance between the functions of university teaching staff in favor of scientific production, threatening academic integrity and the quality of university teaching itself (Fernández-Cano, 2021; Laiho et al., 2020).

Concerning the consequences generated by these criteria, the health and sociofamily concerns are particularly predominant among our participants. In particular, the teachers agree that the quest to fulfill the criteria generate high stress and anxiety levels. These symptoms could have various causes, such as the excessive workload required to pass the assessments, the inability or fear of not passing the assessments or the fear of being socially rejected if the criteria are not met (Dugas et al., 2020; Saura and Bolívar, 2019; Ursin et al., 2020). However, these feelings can also be generated by the internal struggle resulting from the clash between what should be done (adopting the values of the new university, prioritizing research and increasing production, neglecting teaching) and what teaching staff believe should be done (personal or traditional values and beliefs) (Jiménez, 2019; Tülübass and Göktürk, 2020).

Regarding the sociofamily consequences, our participants emphasize the sacrifices made in their personal lives (family, leisure, free time) as family commitments are perceived as an element that slows down or hinders professional advancement. Specifically, multiple studies highlight parenthood as one of the most detrimental elements for academic careers (Harris
At the same time, Andalusian teachers mention how dedication to leisure and free time generates feelings of remorse and guilt for abandoning professional work. These feelings lead to a “voluntary” increase in the workload, resulting in an increase in the health and socio-family consequences described above (Bolívar and Mula-Falcón, 2022).

In addition, it is important to note that the teachers in our study also understand that the evaluation criteria do not contribute to peer cooperation. On the contrary, the new higher education system fosters competitive processes that generate division among colleagues, undermining professional collaboration and thus increasing individualization (Saura and Bolívar, 2019; Djerasimovic and Villani, 2019).

Despite the aforementioned social, professional and health impacts, the search for employment outside academia did not emerge as one of the main consequences in the sample analyzed. This could be due either to the “unconditional passion” for academic work (Bozzon et al., 2017, p. 335) or to the continued hope for a better professional future once stability is achieved (Acker and Webber, 2016).

Finally, our analyses revealed differences in perceptions according to gender, professional category and academic discipline. Concerning the gender variable, women showed higher scores concerning the consequences generated by the evaluation criteria. This finding agrees with a large body of evidence suggesting that the traditional gender roles of the family (which associate women with caring roles) and of the university culture itself (predominantly masculinized) pose a series of added challenges for female academics that make it difficult for them to fulfill the assessment criteria, accentuating the negative consequences of such a system (El-Far et al., 2020; Lipton, 2020; Hortal and Tang, 2023).

Analysis of the professional category variable suggests the existence of differences between the most junior and senior categories. Specifically, the most notable differences are observed in the junior categories as these include staff who, due to their professional and employment situation (characterized by precarity and instability), need to navigate the assessment criteria to progress to more stable positions in the current university system (Jiménez, 2019; San Fabián, 2020). In contrast, the impact of these criteria on senior faculty is minimal as these academics have professional stability and thus do not need to pass these evaluations to survive. However, Huang et al. (2016) add that these differences may also be because, to some extent, the more senior staff because of their age and background have been directly involved in creating the evaluation processes and criteria. Hence, they have a more optimistic view of these systems than their junior counterparts.

Regarding academic discipline, we observed differences between the sciences, arts and humanities, engineering and architecture and the branch of social sciences. However, in all comparisons, the differences favor the social sciences area, i.e., participants working in this discipline expressed a more negative view of the assessment criteria and their consequences. These results are due to the fact that social sciences is an area that benefits less from the current evaluation criteria as it has a shorter tradition in the publishing arena, as well as fewer channels available for such production (San Fabián, 2020). However, this may also be due to the higher levels of competitiveness in this discipline, as it is one of the most popular choices of discipline in the Spanish context (ANECA, 2022).

6. Conclusions
The changes that were generated in higher education to deliver improvements and promote quality education have encouraged monitoring strategies that negatively affect the work, personal and social lives of the teaching staff and the quality of the system and its services. This situation is beginning to give rise to a university system characterized by a performative society that adopts principles aimed at meeting market demands (Tomicic, 2019).
These principles could jeopardize the future of universities as agents of transformation and social progress and pose a direct threat to two of the SDGs: Goal 8 (decent work and economic growth) and Goal 4 (quality education).

Given this backdrop, the need arises to review national or regional evaluation criteria and the guidelines issued at the European level. Although Europe has no legislative power over the educational designs of its constituent countries, its guidelines are always understood as a common reference framework for the entire European community (Brøgger, 2019). Therefore, a thorough review of the framework that guides the current world of academia is essential. To this end, it will be imperative to redefine expectations of academic performance, promoting clear, realistic and achievable objectives; revise the concept of university quality; and generate fairer evaluation criteria that address all faculty roles and are aligned with the expectations set. This is the only way to improve the professional conditions of university teaching staff and the quality of the higher education system.

Finally, although this study focuses on the Spanish context, the reality described is not unique. The presence of quality assurance measures that have materialized in the promotion of accountability processes has become a reality across all political agendas worldwide (Browning et al., 2017). Therefore, this study provides a national perspective (thus far understudied) that contributes to the international debate on the issue of accountability systems in higher education. Future lines of research should extend the results of this study by delving further into Spanish teachers’ experiences and conducting country-specific comparisons.

References


Corresponding author
Javier Mula-Falcón can be contacted at: javiermf@ugr.es

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com