

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

Academics'
perceptions

97

Javier Mula-Falcón and Katia Caballero

*Department of Didactics and School Organisation, Faculty of Education,
University of Granada, Granada, Spain*

Received 5 May 2023
Revised 11 July 2023
Accepted 21 August 2023

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

© Javier Mula-Falcón and Katia Caballero. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

Findings: This work was funded by FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades/El Profesorado Novel en las Universidades Andaluzas: Identidades académicas, cuantificadas y digitalizadas (B_SEJ-534-UGR20); by the State Research Agency, Spanish Ministry of Science, and Innovation, through the project “The influence of neoliberalism on academic identities and the level of professional satisfaction”—NEOACADEMIC—(PID2019-105631GA-I00/SRA (State Research Agency)/10.13039/501100011033); and by the grants for the support and promotion of research by the University of Granada in the field of equality through the project “Gender inequalities in the new managerialist university and their impact on women academics—DAMA—” (INV-IGU199-2022). This project has also received funding from the Ministry of Universities (Spain) through the University Teacher Training Grants Programme (FPU19/00942).



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.) (Bolívar 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 Bolívar (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; Tülübas 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 (Brøgger, 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 (MENCION 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.

	Population	Sample	%Population	%Sample	%Total
University of Almeria	940	97	5.3	4.4	10.3
University of Cadiz	1,739	240	9.8	11.0	13.8
University of Cordoba	1,442	173	8.2	7.9	12.0
University of Granada	3,706	608	21.0	27.9	16.4
University of Huelva	918	88	5.2	4.0	9.6
University of Jaen	980	149	5.5	6.8	15.2
University of Malaga	2,591	115	14.7	5.3	4.4
University of Pablo de Olavide	1,026	124	5.8	5.7	12.1
University of Seville	4,331	589	24.5	27.0	13.6
Total	17,673	2,183	100	12.4	

Note(s): Data on the total teaching staff of each institution were provided by the General Secretariat for Universities, Research, and Technology of the Andalusian Regional Government

Source(s): Authors' own creation

Table 1. Population and sample distribution by university

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 H_0 for an $X^2 = 196.34$ (Relevance) and an $X^2 = 125.30$ (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,

	1	2	3	4	5	6	7	<i>X (SD)</i>	<i>Mdn</i>
	<i>R (%)</i>	<i>R (%)</i>	<i>R (%)</i>	<i>R (%)</i>	<i>R (%)</i>	<i>R (%)</i>	<i>R (%)</i>		
16. The evaluation criteria for accreditation and six-year periods determine the way I direct my academic activity	377 (17.3)	200 (9.2)	157 (7.2)	267 (12.2)	318 (14.6)	391 (17.9)	473 (21.7)	4.38 (2.159)	5
17. The evaluation criteria for accreditation and six-year periods adequately assess my professional quality	583 (26.7)	371 (17.0)	367 (16.8)	346 (15.8)	268 (12.3)	167 (7.7)	81 (3.7)	3.08 (1.779)	3
18. The evaluation criteria have a positive impact on the improvement of my academic activity	687 (31.5)	383 (18.0)	348 (15.9)	322 (14.8)	230 (10.5)	133 (6.1)	70 (3.2)	2.86 (1.747)	3
19. The assessment criteria for accreditation and six-year periods favor cooperation between colleagues	859 (39.3)	407 (18.6)	291 (13.3)	282 (12.9)	164 (7.5)	134 (6.1)	46 (2.1)	2.57 (1.703)	2
24. The teaching evaluation criteria favor my involvement in improving my students' learning	563 (25.8)	378 (17.3)	283 (13.0)	342 (15.7)	252 (11.5)	253 (11.6)	112 (5.1)	3.25 (1.904)	3
25. The priority given to research merits in the evaluation process has a negative impact on my teaching activity	417 (19.1)	225 (10.3)	150 (6.9)	260 (11.9)	253 (11.6)	378 (17.3)	500 (22.9)	4.30 (2.232)	5
29. The research evaluation criteria have a positive impact on the quality of my research activity	495 (22.7)	309 (14.2)	289 (13.2)	374 (17.1)	320 (14.7)	274 (12.6)	122 (5.6)	3.47 (1.900)	3
30. The research evaluation criteria incentivize my academic output	409 (18.7)	231 (10.6)	203 (9.3)	302 (13.8)	340 (15.6)	419 (19.2)	279 (12.8)	4.06 (2.068)	4
31. The research evaluation criteria contribute to the rigor of my research processes	553 (25.3)	319 (14.6)	284 (13.0)	296 (13.6)	296 (13.6)	271 (12.4)	164 (7.5)	3.43 (1.994)	3

Source(s): Authors' own creation

Table 2.
Scores per item on
Dimension 1

Table 3.
Scores per item on
Dimension 2

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

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.

	Sex	Professional category	Scientific discipline
D1	0.575	0.05*	<0.001*
D2	0.00*	0.000*	<0.001*

Note(s): The significance level is 0.05

*Reject the null hypothesis

Source(s): Author's own creation

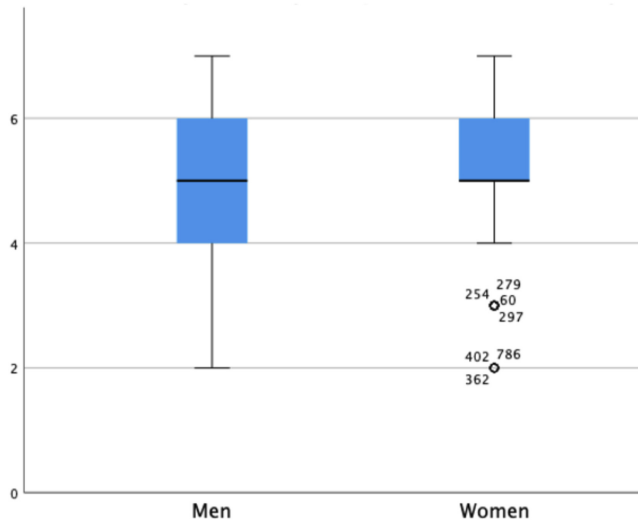
Table 4.
Degree of statistical
significance according
to variable and
dimension

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.



Source(s): Author's own creation

Figure 1.
Joint box plots
Dimension 2 for the
gender variable

Sample/sample	D1	D2
Beginner/intermediate	1,000	0.030*
Advanced/intermediate	0.012*	0.000*
Advanced/beginner	0.052	0.000*

Table 5.
Pairwise comparisons
for the professional
category variable

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): Authors' own creation

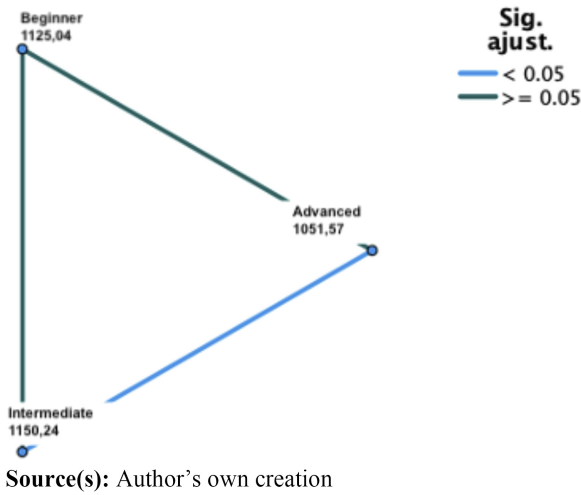


Figure 2. Pairwise comparisons for the professional category variable of Dimension 1

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

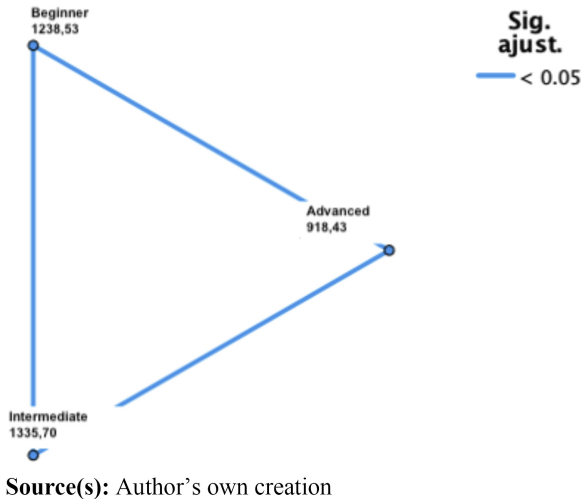


Figure 3. Pairwise comparisons for the professional category variable of Dimension 2

Sample-sample	D1	D2
Sciences-arts and humanities	1.000	1.000
Science-engineering and architecture	1.000	1.000
Sciences-health sciences	0.708	0.075
Sciences-cc. Sciences-social and legal sciences	0.003*	0.001*
Arts and humanities-engineering and architecture	1.000	1.000
Arts and humanities-health sciences	0.901	0.260
Arts and humanities-social and legal sciences social and legal sciences	0.005*	0.011*
Engineering and architecture-health sciences	1.000	1.000
Engineering and architecture-social and legal sciences	0.026*	0.164
Health sciences-social and legal sciences	1.000	1.000

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

Table 6.
Pairwise comparisons
for academic discipline

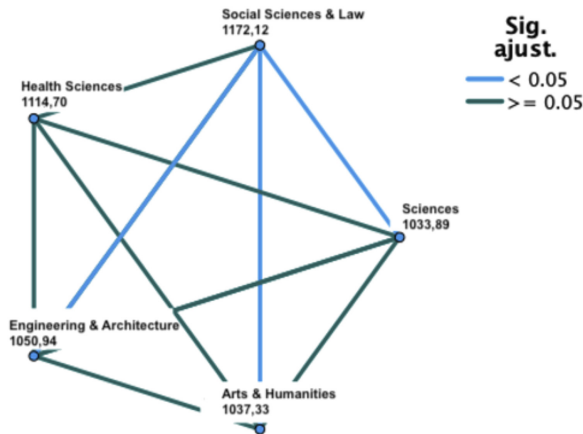


Figure 4.
Pairwise comparisons
for academic discipline
(Dimension 1)

Source(s): Author's own creation

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

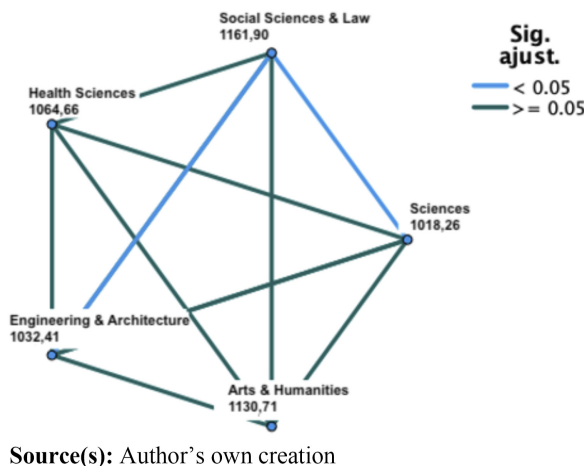


Figure 5.
Pairwise comparisons
for academic discipline
(Dimension 2)

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; Tülübbass 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

et al., 2019; Hortal and Tang, 2023). 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

- Acker, S. and Webber, M. (2017), "Made to measure: early career academics in the Canadian university workplace", *Higher Education Research and Development*, Vol. 36 No. 3, pp. 541-554, doi: [10.1080/07294360.2017.1288704](https://doi.org/10.1080/07294360.2017.1288704).
- ANECA (2022), *Evaluación, Principios, Orientaciones Y Procedimientos*, Organismo Autónomo adscrito al Ministerio de Universidades, available at: https://biblioguias.unex.es/ld.php?content_id=33893908
- Bolívar, A. and Mula-Falcón, J. (2022), "La otra cara de la evaluación del profesorado universitario: investigación vs. docencia", *Revista E-Psi*, Vol. 11 No. 1, pp. 112-129.
- Bozzon, R., Murgia, A. and Poggio, B. (2017), "Work-life interferences in the early stages of academic careers: the case of precarious researchers in Italy", *European Educational Research Journal*, Vol. 16 Nos 2-3, pp. 332-351.
- Brøgger, K. (2019), *Governing through Standards: the Faceless Masters of Higher Education*, Springer, Cham, doi: [10.1007/978-3-030-00886-4](https://doi.org/10.1007/978-3-030-00886-4).
- Browning, L., Thompson, K. and Dawson, D. (2017), "From early career researcher to research leader: survival of the fittest?", *Journal of Higher Education Policy and Management*, Vol. 39 No. 4, pp. 361-377, doi: [10.1080/1360080X.2017.1330814](https://doi.org/10.1080/1360080X.2017.1330814).
- Djerasimovic, S. and Villani, M. (2019), "Constructing academic identity in the European higher education space: experiences of early career educational researchers", *European Educational Research Journal*, Vol. 19 No. 3, pp. 247-268, doi: [10.1177/1474904119867186](https://doi.org/10.1177/1474904119867186).
- Dougherty, K. and Natow, R. (2019), "Analysing neoliberalism in theory and practice: the case of performance-based funding for higher education", *Centre for Global Higher Education Working Paper Series*, available at: <https://www.researchcghe.org/perch/resources/publications/cghe-working-paper-44.pdf>
- Dugas, D., Stich, A., Harris, L. and Summers, K. (2020), "I'm being pulled in too many different directions': academic identity tensions at regional public universities in challenging economic times", *Studies in Higher Education*, Vol. 45 No. 2, pp. 312-326.

- El-Far, M., Sabella, A. and Vershinina, N. (2020), “Stuck in the middle of what?: the pursuit of academic careers by mothers and non-mothers in higher education institutions in occupied Palestine”, *Higher Education*, Vol. 81, pp. 685-705, doi: [10.1007/s10734-020-00568-5](https://doi.org/10.1007/s10734-020-00568-5).
- Fernandez-Cano, A. (2021), “Letter to the Editor: publish, publish . . . cursed!”, *Scientometrics*, Vol. 126, pp. 3673-3682, doi: [10.1007/s11192-020-03833-7](https://doi.org/10.1007/s11192-020-03833-7).
- Gunn, A. (2018), “The UK teaching excellence framework (TEF): the development of a new transparency tool”, in Curaj, A., Deca, L. and Pricopie, R. (Eds), *European Higher Education Area: the Impact of Past and Future Policies*, Springer, pp. 505-526.
- Harris, C., Myers, B. and Ravenswood, K. (2019), “Academic careers and parenting: identity, performance and surveillance”, *Studies in Higher Education*, Vol. 44 No. 4, pp. 708-718, doi: [10.1080/03075079.2017.1396584](https://doi.org/10.1080/03075079.2017.1396584).
- Harvey, L. (2006), “Impact of quality assurance: overview of a discussion between representatives of external quality assurance agencies”, *Quality in Higher Education*, Vol. 12 No. 3, pp. 287-290.
- Harvey, L. (2020), “Research fraud: a long-term problem exacerbated by the clamour for research grants”, *Quality in Higher Education*, Vol. 26 No. 3, pp. 243-261, doi: [10.1080/13538322.2020.1820126](https://doi.org/10.1080/13538322.2020.1820126).
- Hazellkorn, E. (2018), “The accountability and transparency agenda: emerging issues in the global era”, in Curaj, A., Deca, L. and Pricopie, R. (Eds), *European Higher Education Area: the Impact of Past and Future Policies*, Springer, pp. 423-440.
- Hortal, H. and Tang, L. (2023), “Male and female academics’ gendered perceptions of academic work and career progression in China”, *Higher Education Quarterly*, Vol. 22 No. 3, pp. 515-536, doi: [10.1111/hequ.12419](https://doi.org/10.1111/hequ.12419).
- Huang, Y., Pang, S. and Yu, S. (2016), “Academic identities and university faculty responses to new managerialist reforms: experiences from China”, *Studies in Higher Education*, Vol. 43 No. 1, pp. 154-172, doi: [10.1080/03075079.2016.1157860](https://doi.org/10.1080/03075079.2016.1157860).
- Huisman, J. (2018), “Accountability in higher education”, in Nuno, P. and Shin, J. (Eds), *Encyclopedia of International Higher Education Systems and Institutions*, Springer, London, pp. 1-5.
- Huisman, J., Boer, H., Dill, D. and Souto-Otero, M. (2015), *The Palgrave International Handbook of Higher Education Policy and Governance*, Palgrave.
- Jiménez, M. (2019), “Identidad académica: una franquicia en construcción”, *Educar*, Vol. 55 No. 2, pp. 543-560, doi: [10.5565/rev/educar.960](https://doi.org/10.5565/rev/educar.960).
- Laiho, A., Jauhiainen, A. and Jauhiainen, A. (2020), “Being a teacher in a managerial university: academic teacher identity”, *Teaching in Higher Education*, Vol. 25 No. 2, pp. 249-266, doi: [10.1080/13562517.2020.1716711](https://doi.org/10.1080/13562517.2020.1716711).
- Lipton, B. (2020), *Academic World in Neoliberal Times*, Palgrave MacMillan, Sydney.
- Ma, L. and Ladisch, M. (2019), “Evaluation complacency or evaluation inertia? A study of evaluative metrics and research practices in Irish universities”, *Research Evaluation*, Vol. 28 No. 3, pp. 209-217, doi: [10.1093/reseval/rvz008](https://doi.org/10.1093/reseval/rvz008).
- Macheridis, N. and Paulsson, N. (2021), “Tracing accountability in higher education”, *Research in Education*, Vol. 110 No. 1, pp. 78-97, doi: [10.1177/0034523721993143](https://doi.org/10.1177/0034523721993143).
- Mula-Falcon, J. and Caballero, K. (2022), “Neoliberalism and its impact on academics: a qualitative review”, *Research in Post-Compulsory Education*, Vol. 27 No. 3, pp. 373-390, doi: [10.1080/13596748.2022.2076053](https://doi.org/10.1080/13596748.2022.2076053).
- Pattaro, A., Moura, P. and Kruijf, A. (2022), “Transparency and accountability in higher education as a response to external stakeholders and rules: a comparison between ThreeCountry-CaseStudies”, in Caperchione, E. and Bianchi, C. (Eds), *Governance and Performance Management in Public Universities*, Springer, pp. 15-49.
- San Fabián, J. (2020), “El reconocimiento de la actividad investigadora universitaria como mecanismo de regulación del mercado académico”, *Revista de Educación de la Universidad de Málaga*, Vol. 1 No. 1, pp. 23-24, Mággenes, doi: [10.24310/mgnmar.v1i1.7208](https://doi.org/10.24310/mgnmar.v1i1.7208).

- Santos, A., Muñoz-Rodríguez, D. and Poveda, M. (2015), "En cuerpo y alma': insatisfacción y precariedad en las condiciones de trabajo del profesorado universitario", *Arxius*, Vol. 32, pp. 13-14.
- Saura, G. and Bolívar, A. (2019), "Sujeto académico neoliberal: cuantificado, digitalizado y bibliometrificado", *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, Vol. 17 No. 4, pp. 9-26, doi: [10.15366/reice2019.17.4.001](https://doi.org/10.15366/reice2019.17.4.001).
- Schindler, L., Puls-Elvidge, D., Welzant, H. and Crawford, L. (2015), "Definitions of quality in higher education: a synthesis of the literature", *Higher Learning Research Communications*, Vol. 5 No. 3, pp. 3-13, doi: [10.18870/hlrc.v5i3.244](https://doi.org/10.18870/hlrc.v5i3.244).
- Serbetar, I. and Sedlar, I. (2016), "Assessing reliability of a multi-dimensional scale by coefficient alpha", *Revija Za Elementarno Izobrazevanje*, Vol. 9 Nos 1/2, p. 189.
- Stensaker, B. (2018), "External quality assurance in higher education", in Shin, J.C. and Teixeira, P. (Eds), *Encyclopedia of International Higher Education Systems and Institutions*, Springer, doi: [10.1007/978-94-017-9553-1_523-1](https://doi.org/10.1007/978-94-017-9553-1_523-1).
- Stensaker, B. and Maassen, P. (2015), "A conceptualization of available trust-building mechanisms for international quality assurance of higher education", *Journal of Higher Education Policy and Management*, Vol. 37 No. 1, pp. 30-40.
- Tasopoulou, K. and Tsiotras, G. (2017), "Benchmarking towards excellence in higher education", *Benchmarking: An International Journal*, Vol. 24 No. 3, pp. 617-634, doi: [10.1108/BJJ-03-2016-0036](https://doi.org/10.1108/BJJ-03-2016-0036).
- Tavakol, M. and Dennick, R. (2011), "Making sense of Cronbach's alpha", *International Journal of Medical Education*, Vol. 2, p. 53.
- Teichler, U., Arimoto, A. and Cummings, W. (2013), *The Changing Academic Profession: Major Findings of a Comparative Survey*, Springer, Kassel.
- Tomicic, A. (2019), "American dream, Humboldtian nightmare: reflections on the remodelled values of a neoliberalized academia", *Policy Futures in Education*, Vol. 17 No. 8, pp. 1057-1077, doi: [10.1177/1478210319834825](https://doi.org/10.1177/1478210319834825).
- Tülübass, T. and Göktürk, S. (2020), "Neoliberal governmentality and performativity culture in higher education: reflections on academic identity", *Research in Educational Administration and Leadership*, Vol. 5 No. 1, pp. 198-232, doi: [10.30828/real/2020.1.6](https://doi.org/10.30828/real/2020.1.6).
- Ursin, J., Vähäsantanen, K., McAlpine, L. and Hökkä, P. (2020), "Emotionally loaded identity and agency in Finnish academic work", *Journal of Further and Higher Education*, Vol. 44 No. 3, pp. 311-325, doi: [10.1080/0309877X.2018.1541971](https://doi.org/10.1080/0309877X.2018.1541971).
- Ylijoki, O., Lyytinen, A. and Marttila, L. (2011), "Different research markets: a disciplinary perspective", *Higher Education*, Vol. 62, pp. 721-740, doi: [10.1007/s10734-011-9414-2](https://doi.org/10.1007/s10734-011-9414-2).
- Zenatta, E., Ponce, T., García, L., Sánchez, C. and Gama, J. (2017), "Diseño del cuestionario: estrategias identitarias de académicos universitarios ante las reformas educativas", *Revista de Psicología*, Vol. 35 No. 2, pp. 703-724, doi: [10.18800/psico.201702.011](https://doi.org/10.18800/psico.201702.011).

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