

Spanish Scientists' Opinion about Science and Researcher Behavior

Dolores Frias-Navarro¹ , Marcos Pascual-Soler² , José Perezgonzalez³ , Héctor Monterde-i-Bort¹  and Juan Pascual-Llobell¹

¹ *Universitat de València (Spain)*

² *ESIC Business & Marketing School (Spain)*

³ *Massey University (New Zealand)*

Abstract. We surveyed 348 Psychology and Education researchers within Spain, on issues such as their perception of a crisis in Science, their confidence in the quality of published results, and the use of questionable research practices (QRP). Their perceptions regarding pressure to publish and academic competition were also collected. The results indicate that a large proportion of the sample of Spanish academics think there is a crisis in Science, mainly due to a lack of economic investment, and doubts the quality of published findings. They also feel strong pressure to publish in high impact factor journals and a highly competitive work climate.

Received 27 May 2020; Revised 4 November 2020; Accepted 27 December 2020

Keywords: false positives, questionable research practices, replication, scientific integrity, survey

The analysis of questionable research practices is a topic that has aroused considerable interest since the beginning of the 21st century, due to its link to the controversy of the so-called crisis in Science, a controversy that is directly related to the debate on the lack of replication of published findings (Baker, 2016; Benjamin et al., 2017; Fanelli, 2018; Frias-Navarro et al., 2020; Ioannidis, 2005, 2019; Kerr, 1998).

The debate on the crisis in Science itself probably dates back to classical discussions related to publication bias, the use and abuse of statistical significance tests, and the lack of philosophical understanding of the statistical inference process (see revisions by Giner-Sorolla, 2018; Llobell et al., 2000); Monterde i Bort et al., 2006). Indeed, Altman (1994) states that “we need less research, better research, and research done for the right reasons”, and points out that it is common for researchers to use the wrong techniques, employ the right techniques in the wrong way, misinterpret results, report results selectively, cite the literature selectively, and express conclusions that are not justified by the

findings. Consequently, the literature is full of articles with methodological weaknesses, inappropriate designs, and incorrect analytical methods. Although time has passed, the issues Altman mentioned are still present in the conduct of today's researchers, a problem that contaminates the foundations of Science as a source of rigor and quality.

In addition, lack of training in research design and statistical analysis is a key element that can lead researchers to make irresponsible decisions (Altman, 1994; Frias-Navarro et al., 2020). Determining why the researcher carries out questionable research practices involves taking into account a set of variables linked to the researcher (personality trait, dysfunctional personality, moral attitudes...), the social and institutional context of scientific practice (system of awarding funding to research groups, competitiveness, pressure to publish, reinforcement mechanisms to promote one's work, emphasis on publishing statistically significant results...), and the methodological training received. In short, this is a multi-causal problem (Bouter, 2015).

Questionable research practices (QRPs) are those which, consciously or unconsciously, “massage” the

Correspondence concerning this article should be addressed to Dolores Frias-Navarro. Universitat de València. Departament de Metodologia de les Ciències del Comportament. Valencia (Spain).

E-mail: M.Dolores.Frias@uv.es

We thank to all the participants who answered the survey.

Conflicts of Interest: None.

Funding Statement: This work was supported by the University of Valencia (Spain) (Grant Number UV-INV-AE17-698616).

How to cite this article:

Frias-Navarro, D., Pascual-Soler, M., Perezgonzalez, J., Monterde-i-Bort, H., & Pascual-Llobell, J. (2021). Spanish scientists' opinion about science and researcher behavior. *The Spanish Journal of Psychology*, 24, e7. Doi:10.1017/SJP.2020.59

attractiveness of a finding, increasing the prospects of a scientific publication and future citations (John et al., 2012; Simmons et al., 2011). Given that researchers have a certain degree of flexibility throughout the research design process, some of these decisions may be directed at making the finding match the researcher's wish (e.g., a statistically significant result in support of a hypothesis or a non-significant one supporting the assumptions of a statistical procedure), thus increasing the likelihood that a paper with this result will be published (Blanco et al., 2017; Matthes et al., 2015). Such practices may occur with or without intent to deceive (Banks, O'Boyle, et al., 2016; Banks, Rogelberg, et al., 2016). Nor are all practices occurring at the time a researcher analyses data questionable. Indeed, Banks, Rogelberg, et al. (2016) differentiate between practices that pose no problems, practices that imply suboptimal usage but are not overly problematic, and QRPs that pose a serious threat to the inferences made based on the results reported.

Questionable research practices motivate researchers to make decisions designed to achieve desirable results in their studies, leading to *p*-hacking (forcing the data until they are statistically significant), harking (elaborating hypotheses after the findings are known), sharking (removing hypotheses after the findings are known), and cherry-picking (e.g., reporting only findings that confirm the researcher's hypotheses, or fit indices in structural equation modeling) (Hollenbeck & Wright, 2017; Rubin, 2017). QRPs differ from scientific fraud insofar QRPs do not fabricate or falsify data for the purpose of publishing fictitious results (Wells & Farthing, 2019). They are questionable simply because they distort the data in order to support the researcher's goals. In brief, they are "design, analytic, or reporting practices that have been questioned because of the potential for the practice to be employed with the purpose of presenting biased evidence in favor of an assertion" (Banks, O'Boyle, et al., 2016, p. 7).

On the other hand, some actions artificially increase the scientist's production or impact on the literature, such as self-plagiarism, "salami slicing" (segmenting the results of a study in order to produce several publications); honorary authorships (including those based on the hope of reciprocal authorship in future publications), 'ghost' authorship, and excessive use of self-citations oftentimes not relevant to the research (Ding et al., 2020; Hollenbeck & Wright, 2017).

Fanelli's (2009) systematic review of scientific research misconduct presents prevalence data obtained from 18 surveys with international participants, primarily from the biomedical field. The results indicate that up to 14% of researchers believe that scientists fabricate or falsify data, and up to 34% admit that they have performed some questionable research practices, including

making changes in design or results in response to pressures from a funding source. In the case of dishonest conduct by colleagues, up to 72% of the respondents knew of a colleague who had carried out such behaviors. Regarding authorship, Kennedy and Barnsteiner (2014) identify authorship problems in nursing journals, noting that 42% of the articles had honorary authors, and 27% had ghost authorships.

In conclusion, results of meta-research studies and knowledge about researchers' opinions of these questionable research practices are fundamental in addressing this kind of practices. Our study continues this line of work. In fact, this is the first study on this topic carried out in Spain with academic participants from the fields of Psychology and Education. Our main aim is to uncover the perception of questionable research practices by Spanish academics.

Method

Participants

The final sample is a convenience sample of 348 academics from Psychology and Education, with more women (53.4%) than men, between 23 and 69 years ($M = 46.8$, $SD = 10.6$, $Mode = 52$, $Median = 48$). They work mostly at a public university (91.7%) and in a permanent position (58.3%), with work experience ranging between months to 45 years ($M = 16.0$, $SD = 10.75$, $Mode = 10$, $Median = 15$). All respondents identified themselves as Spaniards.

The sample's researcher profile is mostly doctors (85.6%), with moderate participation in the social dissemination of scientific results (60.6%) and publication in indexed journals—Web of Science, Journal Citation Reports (54.9%), as well as acting as peer-reviewers (69.5%), but with little leadership in publicly funded projects (88.5%) or editorship responsibilities, most being neither journal editor (85.1%) nor members or editorial teams (59.9%). As per their research itself, it is mostly of a quantitative orientation (57% of 242 participants), sometimes, but not always, exploring novel hypotheses (61.2% of 242 participants)¹.

Instruments

We collated information using a questionnaire structured into eight sections:

- (i.) Sociodemographic variables.
- (ii.) Perception of a current crisis in Science.
- (iii.) Quality of academic research syllabus.
- (iv.) Confidence in the quality of published scientific results and researcher's ethical behavior.

¹Descriptions are available open access on <https://osf.io/kgvq8>

- (v.) Perception of questionable research practices.
- (vi.) Opinion regarding the publication of statistically significant results, confidence in research conclusions, and fallacies in the interpretation of statistically significant results.
- (vii.) Attitudes and beliefs regarding replication studies.
- (viii.) Perception of pressure to publish and competition in academic contexts².

In regards to statistical fallacies, we asked about three particular fallacies. The “effect size fallacy” presumes that statistical significance informs about the size of the effect, so that small *p*-values equal large effects (Gliner et al., 2001; Kline, 2013). The “clinical or practical significance fallacy” presumes that statistical significance signals clinical or practical significance. And the “finding utility fallacy” presumes that statistical significance signals the usability of the results (Gliner et al., 2001, 2002; Kirk, 1996; Kline, 2013).

Procedure

Participants were canvassed among academic staff listed on the web pages of Psychology and Education departments of Spanish universities, controlling for duplicate entries.

Between May 28, 2018 and June 11, 2018, 3,402 researchers were randomly selected and invited to participate in the study via their publicly available email. A first email notified them of the oncoming survey, including instructions and research objectives. A later email provided the link to the survey, managed via a Computer Assisted Web Interviewing (CAWI) system. A final reminder seven days later was sent to participants who had not accessed the survey in the interim.

A total of 545 surveys were completed (16.02%). The main retention criterion was for participants to have answered all survey items but for three optional questions. 348 participants fulfilled the required criterion; effectively lowering the response rate to 10.23% (242 participants also answered the optional questions).

Analysis were carried out with IBM's SPSS v. 26 for Windows.

Results

Perception of Crisis in Science

63.5% of the sample ($n = 221$) perceives Science to be in crisis. Participants who perceived a crisis had the opportunity to provide their opinion on its causes as an open-ended response. Content analysis of the 144 responses provided by 100 participants ($k = 144$, $n = 100$) on the

Table 1. Open-ended Arguments to Explain the Crisis in Science ($n = 100$)

Reasons for the crisis in Science	<i>k</i>
Lack of economic investment	51
Greater search for quantity	20
Monetary interest in knowledge	8
Lack of institutional support	8
Lack of diffusion/dissemination	4
Too many teaching duties	4
Pressure to publish to improve CV	4
Difficulty of publishing	4
Lack of incentives	3
There are no multidisciplinary teams	3
Instability of the academic and research career (difficult promotion)	3
Lack of training	3
High competitiveness	3
Theoretical models are not interesting	3
Arguments only expressed by one participant	23

Note. Questions: “Do you currently think there is a crisis in Science” and “If your answer to the previous question was ‘Yes’, please indicate why you think this crisis in Science exists”.

main causes attributed to such crisis resulted in two main causes identified: The lack of economic investment ($k = 51$) and an emphasis on quantity of publications over quality ($k = 20$). Overall, for this subset of participants the crisis is perceived as something exogenous to Science, rather than intrinsic to researchers' individual behaviors or organizational and/or social factors (see Table 1).

Research Topics Addressed in University Research Syllabus

Most participants (51.2%) agree that research ethics is explicitly addressed in the course contents of Psychology and Education curricula, followed by meta-analysis, confidence intervals, effect sizes, and scientific misconduct (over 1/3 of participants agree). On the other hand, more than 90% of participants claim no to receive explicit teaching on problems related to the terms *p*-hacking, harking, cherry picking, and sharking, all questionable research practices (however, some 10% to 18% of respondents may be aware of those topics from elsewhere, which results in some 75% to 87% of respondents being unaware of those particular problems) (Figure 1).

These results, however, may be due to respondents being unaware of the English nomenclature, as when asked about the practices in a more narrative manner (e.g., see Figure 2), responses seem to indicate they are more aware of such practices than otherwise claimed.

²Survey, data, and coding are available open access on <https://osf.io/kgvq8>

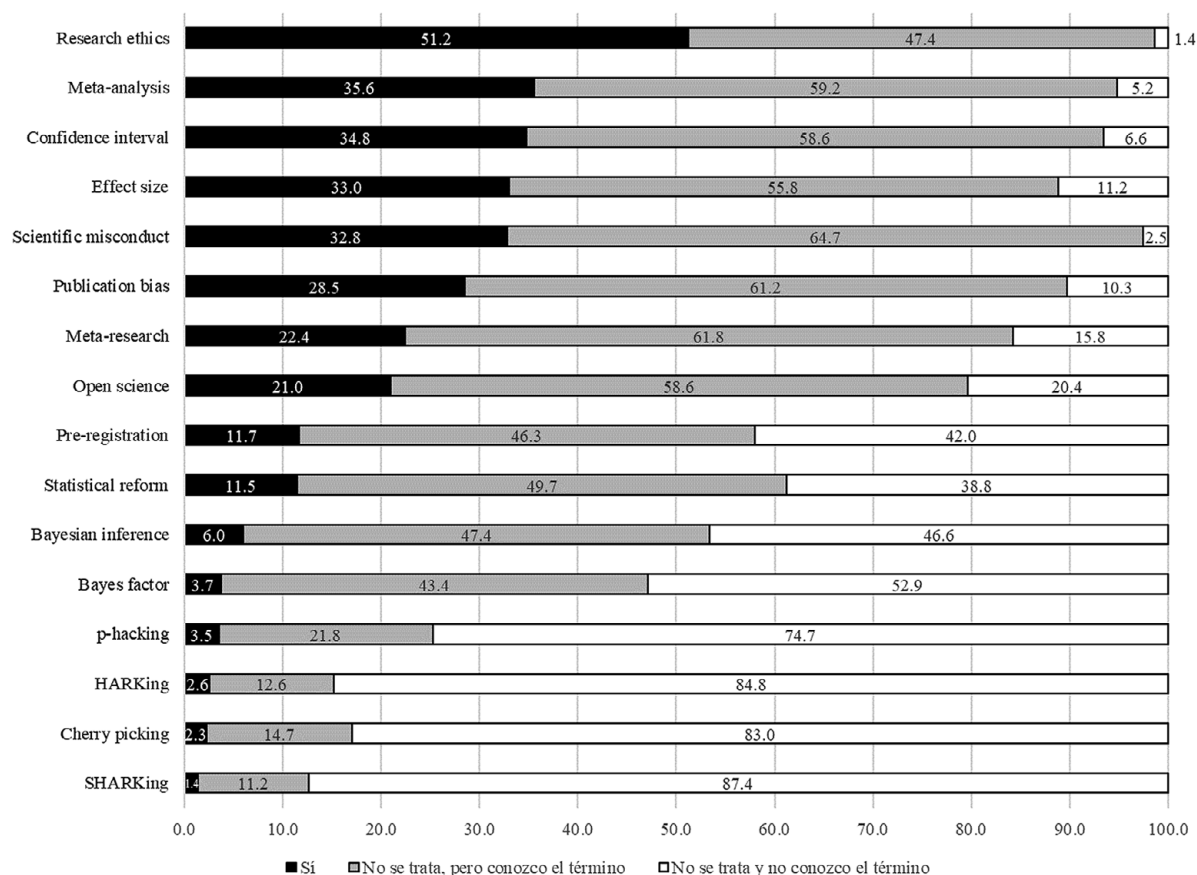


Figure 1. Research Questions Addressed in Course Syllabus

Note. Question: "Does your course program include any topics or do your classes deal with ...".

Confidence in the Quality of Published Findings and in Researchers' Ethics

62.9% of participants express doubts regarding the quality of peer-reviews and 66.7% have doubts about the absence of errors in published studies (Table 2).

As for fraudulent behavior, participants have fewer doubts when they assess their own scientific integrity, that of their own team members and PhD students, and that of other researchers in their own institution. They have greater doubts when assessing the behavior of undergraduate students, followed by graduate students and researchers from other institutions (Table 3).

Questionable Research Practices

The study of fraudulent behavior indicates that only 5.8% of the sample strongly believes that there is fraud in Science. However, it should be noted that 30% indicate that there 'might' be fraud, and that 64.2% categorically state that there is no fraud (Figure 2).

The survey also asked about particular questionable research behaviors, especially those related to

authorship, *p*-hacking, and harking (Figure 2). The practice of listing as co-authors researchers who have not worked on developing or carrying out the study, in exchange for reciprocal co-authorships elsewhere, stands out in first place, with 51.9% of participants strongly agreeing that researchers engage in this type of practice. In second place stands the practice of measuring several variables but only reporting those with statistically significant findings (37%). In third place stands harking (35.9%), that is, rewriting the introduction of the article to hypothesize an otherwise unexpected finding. In fourth and fifth places stand two behaviors that are clear examples of fraud because the study is intentionally manipulated by creating information the author does not have: Citing original studies that have not been read (i.e., fabrication of theoretical information or *p*-literature; 32.8%); and self-citation of articles that have little to do with the topic addressed in the study (falsification of information; 30.3%).

It should be noted that the less extreme 'possibly' response was the option most frequently chosen for most questionable practices, except for those regarding

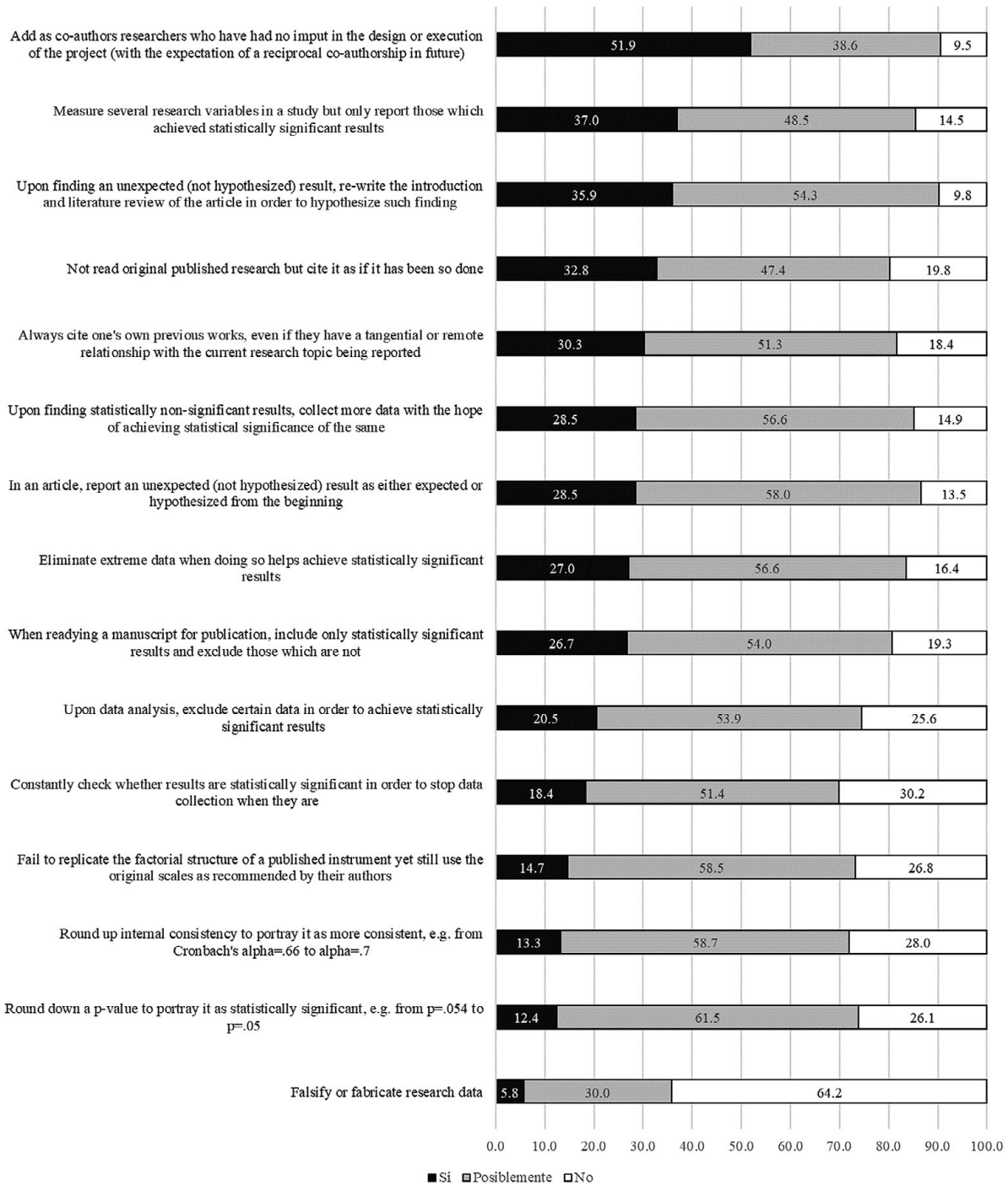


Figure 2. Questionable Research Practices

Note. Question: "Please honestly assess whether you believe that, in research practice, researchers engage in any of the following research behaviors".

co-authorship ('yes' was the most frequent response) and data fraud ('no' was the most frequent response). For example, in regard to the behavior of rounding down the *p*-value to the alpha value (.05), the 'possibly' response is chosen more on this question than on all the other items on the survey (61.5%).

Opinion about Statistically Significant Results

Participants mostly agree (42.5%) or strongly agree (24.1%) that researchers only publish studies when they find statistically significant differences. They also mostly agree (27%) or strongly agree (36.2%) that

Table 2. Confidence in the Quality of Published Results

	1	2	3	4	Mean	SD
1. The tasks of reviewing and correcting manuscripts carried out by journal reviewers reliably guarantee the quality of the scientific results	13.8	49.1	32.2	4.9	2.28	0.76
2. I believe that errors that may have been made in a scientific study are always detected and corrected before being published	18.4	48.3	26.1	7.2	2.22	0.83

Note. Question: "Please rate each of the following issues in relation to your opinion about Science". 1 = Do not agree; 2 = Somewhat agree; 3 = Agree; 4 = Strongly agree.

Table 3. Doubts about Fraud

	1	2	3	4	Mean	SD
1. Your own scientific behavior	80.7	13.8	4.9	0.6	1.25	0.57
2. Your team members	76.7	16.1	6.9	0.3	1.31	0.61
3. Doctoral students	47.1	30.2	18.7	4.0	1.80	0.88
4. Researchers from your institution	47.4	25.3	22.7	4.6	1.84	0.93
5. Researchers from another institution	35.6	31.9	27.9	4.6	2.01	0.91
6. Graduate students	28.2	27.3	35.9	8.6	2.25	0.96
7. Undergraduate students	17.8	22.4	38.2	21.6	2.64	1.01

Note. Question: "To what extent have you doubted the integrity (falsifying, inventing, adding, or removing data) of the research carried out by the following agents?" 1 = Do not agree; 2 = Somewhat agree; 3 = Agree; 4 = Strongly agree.

Table 4. Opinion about Statistically Significant Results

	1	2	3	4	Mean	SD
1. Many researchers only publish studies where statistically significant results are obtained.	6.9	26.5	42.5	24.1	2.84	0.87
2. Scientific journals are not interested in publishing statistically non-significant results.	15.5	21.3	27.0	36.2	2.84	1.08
3. Due to the time required to produce a publication, it would not be a priority to publish findings with statistically non-significant results.	33.6	29.6	21.6	15.2	2.18	1.06
4. When I read an article, I have more confidence in the quality of the study if the results are statistically significant.	32.2	30.2	26.1	11.5	2.17	1.01
5. A scientific conclusion (for example, whether one treatment is better than another) must be based on whether the p -value is or is not statistically significant.	26.4	41.0	22.8	9.8	2.16	0.93
6. Obtaining a statistically non-significant result would be a criterion for not continuing to investigate these variables.	52.6	29.9	12.6	4.9	1.70	0.87
<i>Statistical fallacies</i>						
1. A statistically significant result is an important result	21.1	43.5	22.0	13.4	2.28	0.95
2. The value of $p < .05$ confirms that the finding will be useful to the scientific community.	29.7	48.0	17.4	4.9	1.98	0.82
3. The value of $p = .001$ confirms that the effect size was large.	49.2	33.3	11.8	5.7	1.74	0.88

Note. 1 = Do not agree; 2 = Somewhat agree; 3 = Agree; 4 = Strongly agree.

journals are not interested in publishing statistically non-significant results. Yet they are less agreeable with statistical significance (or lack of) determining when to stop research, the conclusions reached, the level of

confidence on the quality of the underlying research, or publication prospects (Table 4).

Regarding statistical fallacies linked to the interpretation of the p -value, only 32 academics (9.2%) "strongly

Table 5. Attitudes and Beliefs about Replication Studies and the Novelty of Hypotheses

	1	2	3	4	Mean	SD
1. When the results of different investigations are contradictory, it is necessary to carry out a replication study on the same topic.	3.4	11.5	32.5	52.6	3.34	0.81
2. Replication studies are necessary in order to advance science.	2.0	15.5	34.8	47.7	3.28	0.80
3. If the results on a certain topic obtained by different research teams are unanimous, replication studies are unnecessary.	27.6	39.9	23.3	9.2	2.14	0.93
4. Replication studies only make sense when no statistically significant differences are detected in the original study.	64.1	23.3	10.6	2.0	1.51	0.76
5. It is not necessary to replicate a study if statistically significant effects were detected in the original study.	66.1	23.3	8.3	2.3	1.47	0.75
<i>Studies with novel hypotheses</i>						
1. The main objective of scientific journals is to publish novel findings.	17.2	31.9	34.8	16.1	2.50	0.96
2. Science advances more with studies that propose novel hypotheses than with studies that replicate previous research	22.4	37.6	31.1	8.9	2.26	0.91

Note. 1 = Do not agree; 2 = Somewhat agree; 3 = Agree; 4 = Strongly agree.

Table 6. Perception of Pressure to Publish and Academic Competition

	Mean (SD)	Mode	Median	Minimum	Maximum
Perception of pressure to publish in high-impact journals	8.59 (2.01)	10	9	0	10
Perception of competitiveness in university academic activity	8.78 (1.62)	10	9	1	10

Note. Likert scale with 11 anchors, running from 0 = No pressure/no competition to 10 = Very strong pressure/competition.

disagree" with all three fallacies. Thus, the majority of the sample commit one of the three fallacies (agreeing somewhat to strongly), highlighting the opinion that a statistically significant finding is an important and useful result in practice.

Opinion about Replication Studies

Participants almost unanimously point out that replication studies are necessary for Science to advance (98% agree somewhat to strongly). In addition, they think that replication is necessary when findings from different studies are contradictory (96.6% agree somewhat to strongly) yet unnecessary when findings are unanimous (72.4%). Moreover, most do not agree with linking the need for replication to the positive or negative results of a previous study (Table 5).

With regard to conducting only novel studies (versus replication studies), most participants believe that the main objective of scientific journals is to publish novel findings (82.8% agree somewhat to strongly), and that science advances more with studies that have novel hypotheses than with studies that replicate other research (69.49% agree somewhat to strongly), which correlates with an earlier tendency for most participants to carry out studies with novel hypotheses (see 'Participants' section).

Pressure to Publish and Academic Competition

Finally, participants also report high levels of academic pressure and competition (Table 6).

Discussion

The study of questionable research practices can be framed in the area of scientific integrity and ethics, within a climate of perverse and hyper-competitive incentives, which Edwards and Roy (2017) describe as a corrupt academic culture. Such practices are equally related to problems of statistical comprehension and data interpretation (Badenes-Ribera et al., 2015). As Nosek et al. (2012) point out, the professional success of an academic scientist depends on publication, and publication standards support novel and positive results, thus generating incentives that skew publications and, at the same time, the researcher's conduct. Our results indicate that slightly more than two-thirds of the academics surveyed (63.5%) express doubts about the quality of published findings. The results on the perception of questionable research practices show that some academics (51.9%) are particularly concerned about false authorship because increased competition is coupled with fraud, which inflates the curriculum vitae of someone who might be a rival in Academia.

A surprisingly small percentage of respondents categorically state that the questionable behaviors analyzed do not occur among researchers.

The overall picture we gain from our results is that most respondents believe that researchers only publish statistically significant results (93.1%) and that science advances most when novel hypotheses are proposed (77.6%), that scientific journals are not interested in publishing null results (84.5%) but in publishing novel findings (82.8%), that replication studies are necessary when the published findings are contradictory (96.6%) but less so if the findings in the literature are unanimous (72.4%). In addition, over 50% of the respondents misinterpret the meaning of a statistically significant result and associate it with importance, the usefulness of the finding, and the size of the effect (Krueger & Heck, 2019). And while they may not agree with keeping a statistically non-significant result in the drawer (47.4%), they do not consider it a priority to publish these findings (66.4%). In brief, the majority of respondents say that a scientific conclusion (73.6%) should be based on whether or not the *p*-value is statistically significant, and, as readers, they have more confidence in the quality of the study whenever the results are statistically significant (77.8%).

It should be noted that our research measured academics' perceptions of researchers' conduct in general and not the behaviors themselves. We felt that it was more useful to pose the questions in this way in order to avoid the inherent bias of assessing or drawing attention to the researcher's own questionable research practices. If we observe the results related to doubts about the researcher's integrity and fraud, we can see that the majority of the researchers do not doubt their own conduct (80.7%, although it should be noted that 19.3% doubt their own research ethics to some degree) or those of their collaborators (76.7%), focusing their greatest doubts on the practices of the rest of researchers. However, because we did not ask respondents whether they had engaged in QRPs themselves, their answers may be more of a reflection on researchers' degrees of freedom than on scientific fraud proper. Furthermore, some 64% of respondents did not perceive falsification or fabrication of data as occurring, despite well-known exemplar cases such as Diederik Stapel's (Stroebe et al., 2012). However, it is possible that researchers were responding whether they perceived falsification or fabrication of data as routine procedure in current practice, as opposed to awareness of such practices having occurred in the past (thus, about 64% don't perceive that data falsification or fabrication is common in current practice, irrespective of whether it has occurred in the past or not.)

From an individual point of view, the number of publications influences recruitment decisions, salary,

academic promotion, professional recognition, and the likelihood of obtaining a grant. For universities and departments, the number of publications by their academics is also relevant in their classification in international rankings (Ball, 2005; Nosek et al., 2012). Governmental resources for research funding are much less available than researchers would like, and the criteria for accessing stable work in academia are based almost exclusively on the quantitative metrics of impact factors. Moreover, the researcher's excellence is measured using these same criteria, as can be seen in the public standards of Spanish universities. All this has favored a hyper-competitive academic environment, as reported by the academics who participated in our study. They feel highly pressured to publish following the norms of the criteria mentioned above.

In order to interpret the findings of our research, we believe it is necessary to take into account the context of the academic climate and culture (academic promotion of the scientist) as perceived by the researchers, a perception that has been verified in surveys carried out in other countries (e.g., Abbott et al., 2010; Fanelli, 2010; John et al., 2012). Pressure to publish and high competitiveness are two variables that could largely explain why researchers' behaviors become questionable. In addition, journals and their emphasis on novel and positive results (as opposed to replication studies and null results) encourage these questionable behaviors directed at obtaining results that have a high probability of being published (Fanelli, 2012). This leads to 'adjusting' certain aspects of the design, as well as carrying out other behaviors that may alter and improve the researcher's metrics, favored by the degrees of freedom of the researcher's conduct (Neuliep & Crandall, 1990). Certainly, actions such as pre-registration of research and publication of protocols, along with the promotion of open science and the transparency of the research design process, are essential in order to control certain questionable research practices, but the researcher and their personal needs will always lie behind these actions (Chambers, 2019; Nosek et al., 2012).

The results of our research indicate that in answer to the question "Is there a crisis in Science?", approximately two-thirds of the academics surveyed think there is, and they attribute it mainly to a lack of economic investment, followed by the opinion that the quantity of publications takes precedence over its quality. If it is perceived that there are few economic resources and that the system values quantity more than quality, then the direct consequence perceived by the researcher is to 'publish or perish'. Because this involves publishing a lot, and the perception is that there is a greater chance of publishing new and statistically significant results, the researcher's aim is to carry out research that meets those characteristics. The current

research culture, which has been developing for decades (Melton, 1962; Sterling et al., 1995), must change in order to change the researcher's behavior.

Our findings are in line with Baker's (2016) results on confidence in scientific data, but we found more pessimistic opinions. As Baker points out, the area of research is a variable to take into account because, for example, physicists and chemists tend to show more confidence. The results of Baker's survey (2016) also indicate that more than 60% of respondents believe that pressure to publish and selective reporting are the two main factors behind the crisis in Science and the lack of replication, along with the little research being done for replication purposes.

In light of this situation, we believe that support from institutions and funding agencies is essential and indispensable for changing researchers' behavior, along with journal policies and peer review, which must exercise their criteria by analyzing the validity of the results and the quality of the research design process, ignoring any issues not directly related to the scientific method. Publish or perish cannot be a criterion that justifies the researcher's behavior, but a change in incentives is essential as a motivating element for the scientist looking for a job. Certainly, it is difficult to measure scientific performance, which becomes easier when counting the number of articles published, the impact factor of the journals, the number of citations received, the researcher's h index, or the amount of money received from project grants. However, the quality of scientific work is not related to these numbers. It is essential to assess the quality of the evidence provided by the results, and to do so, it is necessary to read the work and check the key elements that contradict the different dimensions of its validity. The data themselves are not the most important aspect, but rather the procedure through which these data were obtained. And this type of assessment, directly aimed at the quality of the scientific method, is a means of improving the quality of science. It focuses on obtaining reliable and valid results that can be published in journals based on the quality of their contribution to scientific knowledge, regardless of whether the result was positive or not. To carry out these types of actions, checklist tools (CONSORT, STROBE, PRISMA...) are quite useful. They require the user (authors, reviewers, editors, or readers) to have methodological knowledge about all the elements being verified because their content tracks the entire process of the scientific method.

One of the most important limitations of our study is the type of sample used. It is a self-selected sample, thus a self-selection bias among respondents cannot be discarded (Bruton et al., 2020). As Baker (2016) points out, it is likely that the respondents were academics concerned about the quality of scientific findings. Furthermore,

those researchers with more confidence in their own research practices may be the ones responding, in which case the findings may as well underestimate the current rate of QRPs (Fraser et al., 2018). Indeed, Banks, Rogelberg, et al. (2016) point out that one of the more problematic concerns may be the underreporting of QRP engagement.

The low response rate (10.23%) is another limitation because it might affect the representativeness of the sample and, consequently, the generalizability of the results. This response rate is similar to what was obtained by other researchers who used the same data collection system with academics (via email): 7% in Bruton et al. (2020), 10.26 % in Badenes-Ribera et al. (2015), 10.58 % in Badenes-Ribera, et al. (2016), or 15% in Fraser et al. (2018). It is also convenient to point out that the interest was on assessing degree of agreement, thus the "biased" scale used (with anchors 1 = *Do not agree*, 2 = *Somewhat agree*, 3 = *Agree*, 4 = *Strongly agree*) allowed to assess such agreement while, at the same time, to focus on the extremes of the scale when interpreting the results).

The crisis of Science has been studied from different perspectives, including economic (e.g., funding constraints, or curriculum building directed towards tenure), failures of replication and credibility, failures in methodological training, and even the degree of social impact of research findings. Our study pertains to the line of research developed in the past decade that has extensively and profoundly reflected on the crisis in Science, questionable research practices, and the need for researchers' statistical re-education. Our study is the first to measure the opinions of Spanish Psychology and Education academics about researchers' behavior and the quality of scientific results. Our findings reflect on ethical behavior because, as Baker (2016) points out, it is healthy for the scientific community to be aware of the problems that surround publication in order to remedy them and provoke changes in researchers' behavior. We fully agree with the recommendations of Dorothy Bishop (2020) and the need to "understand the mechanisms that maintain bad practices in individual humans" in order to understand "why individual scientists mistake bad science for good, and helping them to resist these errors". Approaches to human cognitive biases are not new. For example, in 1976, Mahoney pointed out reviewers' bias toward their favorite ideas. Thus, advancing knowledge about confirmation bias, the degree of morality attributed to errors of omission and commission, statistical fallacies, and the lack of understanding of the concept of conditional probability, which involves the use of the *p*-value (with a key role in planning statistical power and the analysis and interpretation of the data), can improve research practices (Bishop, 2020).

References

- Abbott, A., Cyranoski, D., Jones, N., Maher, B., Schiermeier, Q., & van Noorden, R. (2010). Metrics: Do metrics matter? *Nature*, 465, 860–862. <http://doi.org/10.1038/465860a>
- Altman, D. (1994). The scandal of poor medical research. *BMJ*, 308, 283–284. <http://doi.org/10.1136/bmj.308.6924.283>
- Badenes-Ribera, L., Frias-Navarro, D., Monterde-i-Bort, H., & Pascual-Soler, M. (2015). Interpretation of the *p*-value. A national survey study in academic psychologists from Spain. *Psicothema*, 27(3), 290–295. <http://doi.org/10.7334/psicothema2014.283>
- Badenes-Ribera, L., Frias-Navarro, D., Pascual-Soler, M., & Monterde-i-Bort, H. (2016). Knowledge level of effect size statistics, confidence intervals and meta-analysis in Spanish academic psychologists. *Psicothema*, 28(4), 448–456. <http://doi.org/10.7334/psicothema2016.24>
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533, 452–454. <http://doi.org/10.1038/533452a>
- Ball, P. (2005). Index aims for fair ranking of scientists. *Nature*, 436, 900. <http://doi.org/10.1038/436900a>
- Banks, G. C., O'Boyle, E. H., Pollack, J. M., White, C. D., Batchelor, J. H., Whelpley, C. E., Abston, K. A., Bennett, A., & Adkins, C. (2016). Questions about questionable research practices in the field of management: A guest commentary. *Journal of Management*, 42, 5–20. <https://doi.org/10.1177/0149206315619011>
- Banks, G. C., Rogelberg, S. G., Woznyj, H. M., Landis, R. S., & Rupp, D. E. (2016). Editorial: Evidence on questionable research practices: The good, the bad, and the ugly. *Journal of Business and Psychology*, 31, 323–338. <https://doi.org/10.1007/s10869-016-9456-7>
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Eferson, C., ... Johnson, V. E. (2017). Redefine statistical significance. *Nature Human Behaviour*, 2, 6–10. <http://doi.org/10.1038/s41562-017-0189-z>
- Bishop, D. V. M. (2020). The psychology of experimental psychologists: Overcoming cognitive constraints to improve research: The 47th Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology*, 73(1), 1–19. <http://doi.org/10.1177/1747021819886519>
- Blanco, F., Perales, J. C., & Vadillo, A. (2017). Can Psychology rescue itself? Incentives, biases, and reproducibility. *Anuari de Psicologia de la Societat Valenciana de Psicologia*, 18, 231–252.
- Bouter, L. M. (2015). Commentary: Perverse incentives or rotten apples? *Accountability in Research*, 22(3), 148–161. <http://doi.org/10.1080/08989621.2014.950253>
- Bruton, S. V., Brown, M., & Sacco, D. F. (2020). Ethical consistency and experience: An attempt to influence researcher attitudes toward questionable research practices through reading prompts. *Journal of Empirical Research on Human Research Ethics*, 15(3), 216–226. <https://doi.org/10.1177/1556264619894435>
- Chambers, C. (2019). What's next for registered reports? [Comment]. *Nature*, 573, 187–189. <http://doi.org/10.1038/d41586-019-02674-6>
- Ding, D., Nguyen, B., Gebel, K., Bauman, A., & Bero, L. (2020). Duplicate and salami publication: A prevalence study of journal policies. *International Journal of Epidemiology*, 49(1), 281–288. <http://doi.org/10.1093/ije/dyz187>
- Edwards, M. A., & Roy, S. (2017). Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. *Environmental Engineering Science*, 34(1), 51–61. <http://doi.org/10.1089/ees.2016.0223>
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE*, 4, Article e5738. <http://doi.org/10.1371/journal.pone.0005738>
- Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States Data. *PLOS ONE*, 5(4), Article e10271. <http://doi.org/10.1371/journal.pone.0010271>
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90, 891–904. <http://doi.org/10.1007/s11192-011-0494-7>
- Fanelli, D. (2018). Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 115(11), 2628–2631. <http://doi.org/10.1073/pnas.1708272114>
- Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. *PLoS ONE*, 13, Article e0200303. <https://doi.org/10.1371/journal.pone.0200303>
- Frias-Navarro, D., Pascual-Llobell, J., Pascual-Soler, M., Perezgonzalez, J., & Berrios-Riquelme, J. (2020). Replication crisis or an opportunity to improve scientific production? *European Journal of Education*, 55, 618–631. <https://doi.org/10.1111/ejed.12417>
- Giner-Sorolla, R. (2018). From crisis of evidence to a “crisis” of relevance? Incentive-based answers for social psychology's perennial relevance worries. *European Review of Social Psychology*, 30(1), 1–38. <http://doi.org/10.1080/10463283.2018.1542902>
- Gliner, J. A., Leech, N. L., & Morgan, G. A. (2002). Problems with null hypothesis significance testing (NHST): What do the textbooks say? *The Journal of Experimental Education*, 71(1), 83–92. <https://doi.org/10.1080/00220970209602058>
- Gliner, J. A., Vaske, J. J., & Morgan, G. A. (2001). Null hypothesis significance testing: Effect size matters. *Human Dimensions of Wildlife*, 6(4), 291–301. <http://doi.org/10.1080/108712001753473966>
- Hollenbeck, J. R., & Wright, P. M. (2017). Harking, sharking, and tharking: Making the case for post hoc analysis of scientific data. *Journal of Management*, 43(1), 5–18. <https://doi.org/10.1177/0149206316679487>
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), Article e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Ioannidis, J. P. A. (2019). What have we (not) learnt from millions of scientific papers with *p* values? *The American Statistician*, 73. (Suppl 1), 20–25. <http://doi.org/10.1080/00031305.2018.1447512>
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with

- incentives for truth-telling. *Psychological Science*, 23(5), 524–532. <http://doi.org/10.1177/0956797611430953>
- Kennedy, M. S., & Barnsteiner, J.** (2014). Honorary and ghost authorship in nursing publications. *Journal of Nursing Scholarship*, 46(6), 416–422. <http://doi.org/10.1111/jnu.12093>
- Kerr, N. L.** (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3), 196–217. http://doi.org/10.1207/s15327957pspr0203_4
- Kirk, R. E.** (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56(5), 746–759. <https://doi.org/10.1177/0013164496056005002>
- Kline, R. B.** (2013). *Beyond significance testing: Statistic reform in the behavioral sciences*. American Psychological Association.
- Krueger, J. I., & Heck, P. R.** (2019). Putting the p-value in its place. *The American Statistician*, 73. (Suppl 1), 122–128. <http://doi.org/10.1080/00031305.2018.1470033>
- Llobell, J. P., Perez, J. F. G., & Navarro, M. D. F.** (2000). Statistical significance and replicability of the data. *Psicothema*, 12 (Suppl. 2), 408–412.
- Mahoney, M. J.** (1976). *Scientist as subject: The psychological imperative*. Ballinger.
- Matthes, J., Marquart, F., Naderer, B., Arendt, F., Schmuck, D., & Adam, K.** (2015). Questionable research practices in experimental communication research: A systematic analysis from 1980 to 2013. *Communication Methods and Measures*, 9(4), 193–207. <http://doi.org/10.1080/19312458.2015.1096334>
- Melton, A. W.** (1962). Editorial. *Journal of Experimental Psychology*, 64(6), 553–557. <https://doi.org/10.1037/h0045549>
- Monterde i Bort, H. M. I., Llobell, J. P., & Navarro, M. D. F.** (2006). Interpretation mistakes in statistical methods: Their importance and some recommendations. *Psicothema*, 18(4), 848–856.
- Neuliep, J. W., & Crandall, R.** (1990). Editorial bias against replication research. *Journal of Social Behavior & Personality*, 5(4), 85–90.
- Nosek, B. A., Spies, J. R., & Motyl, M.** (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7(6), 615–631. <http://doi.org/10.1177/1745691612459058>
- Rubin, M.** (2017). When does HARKing hurt? Identifying when different types of undisclosed post hoc hypothesizing harm scientific progress. *Review of General Psychology*, 21(4), 308–320. <http://doi.org/10.1037/gpr0000128>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U.** (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. <http://doi.org/10.1177/0956797611417632>
- Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J.** (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician*, 49(1), 108–112. <http://doi.org/10.1080/00031305.1995.10476125>
- Stroebe, W., Postmes, T., & Spears, R.** (2012). Scientific misconduct and the myth of self-correction in Science. *Perspectives on Psychological Science*, 7(6), 670–688. <http://doi.org/10.1177/1745691612460687>
- Wells, F., & Farthing, M.** (2019). *Fraud and misconduct in biomedical research* (4th Ed.). Taylor & Francis Group. <https://doi.org/10.1201/9780429073328>