Predatory journals and publishers: Characteristics and impact of academic spam to researchers in educational sciences

Jaume Sureda-Negre, Aina Calvo-Sastre, and Rubén Comas-Forgas

University of Balearic Islands, Institute for Educational Research and Innovation, Palma, Spain

ORCID:
J. Sureda-Negre: 0000-0003-2689-5626
A. Calvo-Sastre: 0000-0001-7109-9165
R. Comas-Forgas: 0000-0002-8885-753X

*Corresponding author: Rubén Comas-Forgas, University of Balearic Islands, Institute for Educational Research and Innovation, Palma, Spain. E-mail: rubencomas@uib.es

Abstract

This study focuses on the phenomenon of presumed predatory scientific publications in the field of Educational Sciences, and the utilization of email by editors to request manuscripts. It examined, using content analysis methods, 210 emails received by three professors of the field of Education, at a Spanish university with different research profiles over a period of 3 months. Through analysis of the unsolicited emails a total of 139 journals and 37 publishers were identified and examined using: (a) the two main predatory journal inventories (Beall’s list and Cabells’ Predatory Reports), and (b) six of the major scientific bibliographic databases. The publishers and their websites were also analyzed, as well as the basic aspects of the emails’ content. The majority of the unsolicited emails were from predatory journals or publishers and half of the article requests did not match the field of the recipient. In addition, it is relevant to note that more than half of the domains of predatory publishers analysed have untrustworthy security levels. The data provided relevant information on the phenomenon of predation in scientific publications in the field of Education and, most importantly, provided evidence for developing training and preventive strategies to tackle it.

Keywords: academic ethics, content analysis, scientific communication, science dissemination, academic writing, educational sciences

INTRODUCTION

Digitisation and online publishing in the mid-1990s facilitated open access, and initiated a process of change in the field of scholarly publishing in which we are still immersed today (European-Commission, 2019). In its present state, commercial publishers are still the main service providers, and, as an unintended consequence of the open access paradigm (Beall, 2012) there is a proliferation of scientific journals that, upon payment by the authors, publish their articles quickly, and with poor or no quality controls, especially those related to the peer review system; these are the predatory journals and publishers. This phenomenon began from 2008 to 2009 (Taylor, 2021) and its penetration in scientific...
communication systems has been increasing annually (Bagues et al., 2019; Perlin et al., 2017; Shen & Björk, 2015). It has been estimated that there were 1,800 such journals in 2010, 8,000 in 2015, ~13,000 in 2020 (Boukacem-Zeghmouri et al., 2020), and over 15,000 as listed in Cabells Predatory Reports in 2021 (Cabells, 2021). Therefore, it is not surprising that this situation has been described as one of the greatest current challenges in scientific communication (Machacek & Srholec, 2021) and, consequently, a serious threat to science (Taylor, 2021).

One of the main identifying elements of predatory publishing is the utilization of indiscriminate requests for manuscripts using unsolicited emails (presumed spam), with persuasive and, in a few cases, misleading messages (Beall, 2016; Callaghan & Nicholson, 2020; Cobey et al., 2019). It should be noted that academic spam has increased alarmingly in recent years, with estimated increases of ~40% from 2013 to 2018 (Scheper & Rammelt, 2019). In this context, it is not surprising that one method to address the issue of predatory publishing has been the analysis of unsolicited emails received by academics with invitations to publish their research. This field has provided an important body of knowledge on the topic in recent years.

Most of these studies have focused on academic spam received by researchers in Biomedical Sciences, a field in which this phenomenon has significantly grown, and with a research community exhibiting greater sensitivity and proactivity in handling it (Cohen et al., 2019); however, social sciences and humanities are not immune to the risks associated with predatory publishing, as illustrated, for example, in the study of Bagues et al. (2019), which demonstrates that 5% of the academic articles in Economics submitted in 2012 as contributions for evaluation of their authors in the Italian research system had been published in journals considered predatory.

However, addressing this issue using the analysis of academic spam sent to researchers in Social Sciences and Humanities is very limited (and unprecedented if restricted to the field of Education). Of the existing studies in these areas of knowledge, the first to note is the study by Wahyudi (2017), based on the precedent study by Brown and Cook (2013), which analysed 25 emails from a semiotic perspective for discerning the generic structure and lexical-grammatical characteristics of these emails. The second to note is the study by Soler and Cooper (2019) that analysed the structure of 58 emails from a sociolinguistic perspective, focusing on their rhetorical movements and lexical-grammatical aspects. A somewhat broader approach was adopted by Lund and Wang (2020) who replicated the study of Clemons et al. (2017) conducted in the field of Medical Sciences, with 98 emails received by the authors themselves.

The obvious scarcity of studies focused on Social Sciences and to reiterate their total absence in the field of Educational Sciences, and the fact that none of them have been conducted in the Spanish-speaking cultural context reinforces the relevance of this study that had the following objectives: (1) identify, classify, and quantify academic spam directed at Education researchers, focusing on proposals for submission of articles to journals; (2) analyse the structure of these messages and characteristics of the selected academic journals, and, finally, (3) identify and analyse the characteristics of the publishers of predatory journals.

### Key points
- Almost 70% of the unsolicited emails received by three Spanish education researchers come from predatory publishers and journals.
- An average of 145 academic spam emails are received annually by participants in the study, less than that reported in other disciplines.
- Predatory journals were more likely to mention their impact factor when sending spam emails, but these were often spurious.
- Spam emails frequently mentioned the speed of publication with predatory journals promising an average peer review time of 1 week.
- The level of security of the domains of the academic predatory publishers is very weak.

### MATERIAL AND METHODS

This article documents a descriptive multi-method study based on two data sets. The first was derived from the content analysis of a convenience sample of unsolicited emails, with proposals for publication in academic journals, received by three researchers from the Department of Education of a Spanish public university from the period of 1 January 2021 to 30 March 2021. Content analysis has previously been utilized in studies with emails in general as the object of study (Smith et al., 2012) and studies on spam (Rich, 2018). The researchers collaborating in the study agreed to collect and forward all unsolicited emails received in the inboxes of their institutional accounts with content concerning invitations or academic proposals (participating in conferences, publishing scientific papers, etc.) during the period of analysis to the work team. It is important to note that instructions given in a study explanatory document sent to the three researchers, and in a video conference meeting held 2 weeks before the start of the fieldwork explicitly requested for the study participants to forward all messages with academic invitations to a dedicated email account. They were also reminded of the need to check their spam folders at least once weekly.

Considering the evidence on the relationship between the recipients’ academic status and number of emails which found that young researchers receive substantially more proposals (Cuschieri & Grech, 2018), the participating researchers were selected based on three academic profiles. The first was that of senior researcher with over 20 years of experience from her doctorate, an extensive publication profile (over 30 articles published...
in high-impact journals), and a high number of citations of her publications (h-index of 21 in Google Scholar). The second profile was that of mid-career researcher represented by a professor with 15 years of experience, 11 years from her doctorate, author of 15 to 20 articles published in high-impact journals, and an h-index of 18 in Google Scholar. Finally, the third profile was that of junior researcher represented by a professor with less than 10 years of experience, 3 years from her doctorate, less than 15 articles published in high-impact journals, and an h-index of 11 in Google Scholar.

The second data set resulted from the analysis of various dimensions and characteristics of the domains of the predatory publishers that sent unsolicited emails to the researchers collaborating in the study.

Data collection and analysis
Unsolicited emails with invitations to publish articles in academic journals
A total of 210 emails were collected and initially classified according to the type of proposal or offer they contained, to focus the study solely on those that explicitly proposed to publish articles in journals (n = 97; 46.1% of the spam received), and to eliminate emails with invitations to participate in conferences (44.7%) from the data sets or other proposals (9%) such as invitations to write a chapter or book, product presentations and promotions and editorial news, and invitation to join the team of reviewers of a journal. Messages written in languages other than English and Spanish were also excluded (four messages were received in Russian). The structure of the emails of the final sample of the study (n = 97) was then analysed according to the following dimensions: (a) message content; (b) references to journal impact factors and indexing; (c) temporal references referring to manuscript submission dates and review periods, and, finally, (d) information on publication fees. All these dimensions were derived from the previous study of Clemons et al. (2017) and Lund and Wang (2020). Next, to determine the potential predatory nature of the journals identified in the emails (n = 139, because some of the 97 emails included several titles), their inclusion or exclusion in prestigious registries (Validated list, VL) and malicious repertoires (Predatory list, PL) was checked following the classification established by Misra et al. (2017). Regarding PL, the Beall’s list (https://beallslist.net) and the most recent and complete Cabell’s predatory reports (Chen, 2019; Silver, 2017a, 2017b) were used. As for the VL, each publication was screened based on six major scholarly repertoires and databases with clear evaluation policies: (1) the master journal list-WoS; (2) ScImago journal rank; (3) the directory of open access journal (DOAJ); (4) MIAR; (5) ERIH plus, and (6) DIALNET. Finally, the publishers were identified, analysed and categorized based on similar previously utilized PL and, with respect to the VL, the indices of: (a) OA Scholarly Publishers Association and (b) Committee on Publication Ethics (COPE).

The emails were downloaded as PDF files and stored by the researchers in a shared work folder. The content analysis of the messages followed a similar process as that conducted by Lund and Wang (2020): two of the researchers authoring the study analysed the emails composing the study sample successively according to the previously mentioned dimensions, while the third researcher had to individually review the double analyses of each email and find a solution in case of any divergence between the two researchers regarding data collection.

The data were transferred to three Excel data matrices: the first had the data of all the messages, the second had the data of the journals and the third had that of the publishers. These matrices were then exported to the statistical analysis software SPSS 21 for the data statistical analysis.

Analysis of the websites of predatory publishers
A list of 33 potentially predatory publishers was obtained from the 97 unsolicited messages regarding scientific journals. The domain of these publishers was identified, and the following aspects were analysed: (a) date of creation of the domain using the TCP WHOIS protocol; (b) country where the domain is registered, also obtained with the TCP WHOIS protocol; and (c) security level of the domain obtained from www.ssllabs.com.

The fieldwork for this second part of the study was performed by the three researchers authoring the study during the months of May and June 2021. The obtained data were exported to an Excel file that was subsequently processed for statistical analysis using the SPSS v21 package.

ANALYSIS AND RESULTS
Predatory versus legitimate journals
During the period of the study, participants received 97 academic spam emails with proposals to publish their work in scientific journals, representing a mean of 2.6 weekly, and, if converted into an annual mean, ~145 messages annually, received by each participating researcher. A few of the 97 emails with invitations to publish in journals were not limited to a single title, but rather sought to promote various journals from a similar publisher, raising the total number of journals identified and analyzed to 139 (Table 1).

In total, 83.4% of the journals correspond to potentially predatory journals, considering as such those with titles, or publishers listed on PL or not listed on either PL or VL. Uniquely 16.5% are present on at least one of the VL utilized. If these data are correlated with the annual mean of academic spam calculated above, it can be estimated that each participating researcher receives a mean of ~120 spam messages annually from predatory journals.

Regarding the 23 journals included in the VL, 6 appear in the master journal list (MJL) and 20 in the ScImago journal rank. Most (5) of those appearing in MJL belong to the group of Emerging Sources, and one appears indexed in the JCR,
Occupying positions in the lower part of its ranking: Q4. Most (n = 17) journals indexed in Scimago Journal Rank also appear in the lower quartile of the said impact index, one in Q1, and one in Q2, both belonging to research areas other than Education.

### Concordance between the researcher profile and the subject matter of the journals

Focusing on the subject matter of the journals, it should be noted that 36% of the messages match the recipients’ fields. Conversely, 30.2% of the messages offer publication opportunities in journals that are not related to the field of research of Education and, therefore, do not match the research profiles of the recipients of the emails.

### Academic spam content

Approximately 90% of the spam is not personalized; hence, these messages are sent indiscriminately, while 10.3% is addressed to a recipient, and identifies the recipient by their first and last name. Approximately 22.6% of the messages are signed by a sender or real person, in most of the cases someone who is introduced as the editor or editorial manager of the journal.

One of the fundamental contents, and which is strongly mentioned in the emails, refers to the impact factors of the journals and their indexing in databases, indexes or repositories. Secondly, there are references and allusions to temporal aspects related to manuscript acceptance and, finally, to the review and publication period of the articles, with considerable differences in the treatment of these three aspects of the emails content, depending on whether the messages are from legitimate or potentially predatory journals.

As far as the impact factor is concerned, ~50% of the messages from potentially predatory journals refer to this, while only 12.2% of the emails from legitimate journals refer to it, with the peculiarity that the reference is very brief and unspecific in more than half of the cases of predatory journals, simply stating a number without further explanation. The messages that specify the type of metric provide data unrelated to that of the main international impact meters, using others, some very confusing, such as: scientific journal impact factor; IBI Factor; I2OR Impact factor; SIS Impact factor; AQCJ; ISRA-JIF; PIF; NAAS SCORE. References to the inclusion of the journals in a scientific database, index or directory appeared in 55% of the emails of potentially predatory journals and in 18.7% of the legitimate journals, of which the most cited were Google Scholar, Copernicus, NASA, ANED, Cross Ref, and Open J Gate.

As noted, references to the time between the receipt of the message and deadline for manuscript submission to the journal is another noteworthy aspect of the content of the messages, and was much more present in messages from predatory journals (53% against 18.7% for legitimate journals), with a mean of 10.4 days, while it was 100 days for legitimate journals. There are also considerable differences in the manuscript review and publication periods: 6.2% of emails from legitimate journals refer to the duration of the review period (60 days on average), while the percentage rises to 62.9% for potentially predatory journals with a mean review period of 7.1 days. The mean number of days for publication from the manuscript acceptance date is 9.7 days in the case of predatory journals, with no references to this data in the emails of legitimate journals.

Another analysed dimension is that of publication costs. Regarding this, eight messages from journals (two of which correspond to legitimate journals) refer to these fees, with a mean price of $155, ranging from $66 to $399.

### Publishers sending academic spam

The emails containing proposals to publish articles in journals revealed 37 academic publishers responsible for publications, with most (62.1%) included in lists of predatory publishers. Four of them (18.8%) are included in VL and 10 (27%) are in the grey area of publishers that do not appear on VL or PL (Table 2).

Access to the websites of those publishers led to a more precise analysis of the publication fees, which, as aforementioned, were explicitly stated in 8 of the 97 emails. The article processing charges found in the analysis of the publishers’ websites range from $34 to $2,779 (Table 2). Several different publishers charge fees that vary according to the level of development of the authors’ country, and, in a few cases, to the type of article. In these situations, solely rates for developed countries have been considered, and the rates for all types of documents have been averaged.
TABLE 2 Characteristics of publishers sending academic spam (n = 37): their inclusion on PL and VL and their fees

<table>
<thead>
<tr>
<th>Characteristics of publishers sending academic spam</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of publishers registered on PL</td>
<td>23 (62.1%)</td>
</tr>
<tr>
<td>Total number of publishers on VL (OA Scholarly Publishing Association or COPE)</td>
<td>4 (18.8%)</td>
</tr>
<tr>
<td>Publishers that do not appear on either VL or PL</td>
<td>10 (27%)</td>
</tr>
</tbody>
</table>

Characteristics of the websites of predatory publishers

The data on the country where the domain of the identified publishers is registered indicates a large number within countries with a high level of scientific research output (Table 3).

Another studied dimension was the age of the publishers’ websites, and these results were very striking: the domains of legitimate publishers were an average of 29.5 years, while this average was only 6 years in the case of predatory journals.

The security level of the websites of the analysed publishers was calculated and established using the portal www.ssllabs.com, and an SSL server test was run for each case. The result indicated that three of the four domains of non-predatory publishers had the maximum rating of A+, while the other obtained an A. Of the domains of the 33 predatory publishers, 45.4% and 33.6% have security levels of A and B, respectively, and 21% have server security levels lower than B. This means that more than half of the domains of predatory publishers have poor and unreliable security levels.

DISCUSSION AND CONCLUSIONS

The results of this study should be interpreted considering several limitations of which the authors are fully aware. The first is the sample utilized, because, although the number of emails does not differ considerably from those utilized in similar studies, conducted in other scientific fields, and published in prestigious journals (Dagens, 2019; Lund & Wang, 2020; Nguyen et al., 2018), a larger sample would have conferred greater consistency and robustness to the results. The second limitation is a consequence of current procedures for classifying journals as predatory. Without entering the debate on the existing lists (Buschman, 2020), the fact is that their utilization can generate false positives and negatives (Da-Silva et al., 2020), requiring some caution when interpreting the data thus obtained; however, these limitations should not undermine our results. Firstly, the results indicate that among academics in the field of Social Sciences, and more specifically among researchers in Education, predatory publishers make extensive use of spam to disseminate their products. However, the number of unsolicited emails (0.4 per day) is lower than in other academic environments and other geographical contexts, which range from 0.2 to 12.5 unsolicited emails per day, with a mean of 4.5 (Da-Silva et al., 2020). Nevertheless, a little more than three-quarters of unsolicited messages from journals inviting academics in the field of Education to publish are regarded as predatory, and this is highly revealing of the magnitude and severity of the phenomenon.

One of the main characteristics of predatory journals is their poor contribution for strengthening the research profile of those who publish in them; therefore, solely 4 of the 139 journals found in the study (one indexed in JCR and three with Q1 and Q2 positions in SCIMAGO/SCOPUS) would lead to publications with a positive assessment in accreditation and evaluation processes of researchers in our country. In addition, the results reveal a high percentage of unsolicited emails proposing publication in journals unrelated to the field of research of the recipients of the emails. These results are in line with researches conducted in the field of Biomedical Sciences (Clemons et al., 2017) and Library Science (Lund & Wang, 2020). Given this low relevance,
and considering that the productivity of academics can be negatively affected (Da-Silva et al., 2020; Wilkinson et al., 2019), the most appropriate approach is to tackle this type of messages send by predatory publishers through unsolicited emails or spam.

As for the implications of our results, it is worth noting the risks that predatory publications and their dissemination using spam pose both individually for each researcher, and collectively for the disciplinary field of Social Sciences. The first of these dangers is evident, on the one hand, by the data on the level of security of their websites that clearly discourages the utilization of these websites to make payments of article processing fees, and, on the other hand, by the limited, if not negative, impact of these publications on the research profile of the authors who publish in them. The second level of risk, the collective risk, emerges when considering that disciplines in the field of Social Sciences and Humanities are founded on more questioned epistemologies (Althaus, 2019), and, therefore, more exposed to the risks posed by predatory publications as disseminating structures to legitimize ideologies (Boukacem-Zeghmouri et al., 2020) or simply to spread falsehoods or misinformation. In this context, the collected data and the results can help to address this challenge by the implementation of awareness campaigns and training initiatives for promoting consciousness on the issue among academics in Education, and, consequently, the refusal to submit manuscripts to these publishers (Frandsen, 2019). In addition to illustrating the deceptive processes utilized by predatory publishers, the collected data also provides a warning list that incorporates the most active predatory journals and publishers in the field of Education, which can be useful to facilitate this training. Journals in educational sciences are represented in the major scientific databases: in SCOPUS, for instance, in 2020 the number of indexed journals was 1,536 what correspond to around 4% of the total of journals indexed in this database, in the case of Web of Science the percentage is around 2% and in the DOAJ have a representation of around 9%. Very similar figures are obtained if the presence of educational science journals in predatory journals’ databases is analysed; for instance if we use data from Cabells’ Predatory Reports we found that the percentage of these journals is around 4.5% of the total. Somehow, the representation of educational sciences journals in legitimate scientific databases and predatory journals lists is quite similar.

Finally, it is important to reflect on the importance of reversing some of the features of the current research culture’s ecosytem that has facilitated the emergence and consolidation of predatory publications. One of these characteristics revolves around the principle of ‘publish or perish’ (Wellcome-Trust, 2020); a culture that constitutes the ideal biotope for the emergence and development of phenomena such as predatory journals and publishers (Cobey et al., 2019; Kurt, 2018). The pressure to publish affects the quality of research (Anderson et al., 2007; Fanelli, 2009; Van-Dalen & Henkens, 2012), and is the breeding ground for scientific misconduct.

Research in Education is too crucial to be left in the hands of greedy and unprofessional open access publishers. The open access model based on charging fees for processing articles turns authors into clients, a role previously played primarily by academic libraries; therefore, authors must ensure that they are prepared to work and succeed in this new role.

Researchers must acquire skills in what might be called ‘scholarly publishing literacy’ (Zhao, 2014), which involves acquiring the ability to identify and reject controversial and deceitful offers from companies related to scientific publications or events. This skill also involves knowing how to identify and submit their work to respected journals, book publishers and conferences in their field or area of expertise.

ACKNOWLEDGEMENT
Authors would like to thank the Spanish Ministry of Science and Innovation, the Spanish State Research Agency and the European Regional Development Fund (ERDF) for their support and funding the project RTI2018-098314-B-I00. Open access publishing facilitated by the University of the Balearic Islands, as part of Wiley agreement via the Spanish National Research Council.

AUTHOR CONTRIBUTIONS

REFERENCES


