
A survey on the use of CrossCheck for detecting plagiarism in journal articles

Yuehong (Helen) ZHANG and Xiaoyan JIA
Journal of Zhejiang University-SCIENCE (A/B/C),
PR China

ABSTRACT: *The purpose of this survey was to investigate journal editors' use of CrossCheck, powered by iThenticate, to detect plagiarism, and their attitude to potential plagiarism once discovered. A 22-question survey was sent to 3,305 recipients, primarily scholarly journal editors from Anglophone countries, and a reduced 10-question version to 607 editors from non-Anglophone countries. The response rate was 5.6%. 42% of all respondents had used CrossCheck in their work. The main findings are as follows: (1) the plagiarism detection tool and its similarity report are extremely useful and effective and can assist editors in screening documents suspected of plagiarism; (2) responses show the journal editors' attitude and level of tolerance towards different kinds of plagiarism in different disciplines; (3) the survey results underscore a clear consensus on editorial standards on plagiarism, but there were small variations between different disciplines and countries, as well as between Anglophone and non-Anglophone. The study also suggests that further work is needed to establish a universal principle and practical approaches to prevent plagiarism and duplicate publication.*



Yuehong
(Helen)
Zhang



Xiaoyan Jia

This survey is part of a research study commissioned by the Committee on Publication Ethics with the aim of developing evidence-based guidance for journal editors on how to deal with different kinds of plagiarism detected through the use of CrossCheck (<http://publicationethics.org/resources/research>).

The two authors contributed equally to this work.
© Yuehong (Helen) ZHANG and Xiaoyan JIA 2012

Introduction

Although there are many dictionary definitions of plagiarism,^{1–3} Hames⁴ provides the stricture that ‘duplication of text or results from other articles or books is clearly unacceptable, either from the work of other authors or from an author’s own work (known as auto- or self-plagiarism)’. Plagiarism or unreasonable levels of copying in journal articles is a growing problem: ‘Not so many years ago, we got one or two alleged cases a year. Now we are getting one or two a month.’⁵ ‘How to stop plagiarism’⁶ is an important topic in academic publication. The first step is for journal editors to use various tools to detect real plagiarism in order to protect authority and originality. In recent years, CrossCheck has become ‘an established part of the editorial process for many journals’.^{7–11} And as of February 2012, 291 publishers together accounting for more than 10,000 journals have become members of CrossCheck.¹²

However, in our experience, CrossCheck is just a useful tool to help the editor find strings of similar text. Most instances of true plagiarism cannot be identified solely by these strings. This led us to question: (1) How do journal editors worldwide use CrossCheck and handle the similarity reports? (2) What are journal editors’ attitudes and tolerance toward types of plagiarism in different disciplines and different countries? (3) What are mainstream views and differences to these problems between editors in native English-speaking countries and non-native English-speaking countries? With this in mind, we undertook a survey as part of a research project funded by the Committee on Publication Ethics¹³ (questionnaire in Appendix 1, available online).

Methodology

Survey design

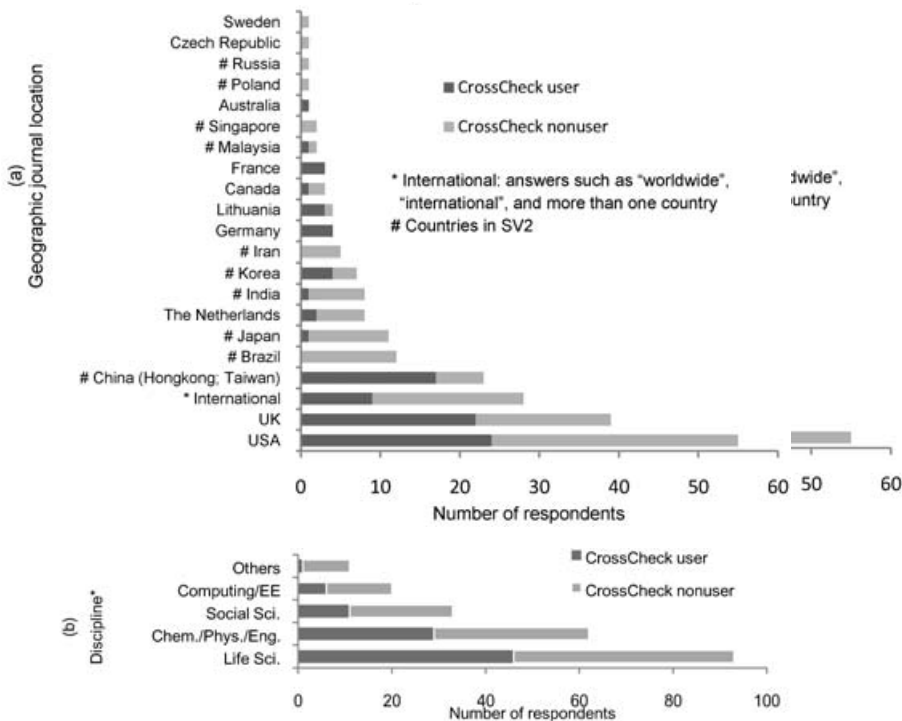
From May to June 2011, an online survey was carried out, using SurveyMonkey (<http://www.surveymonkey.com>). The questionnaires were sent in two separate groups:

1. Survey Version 1 (SV1, the full questionnaire), consisting of 22 questions, was sent to 3,305 recipients mostly from Anglophone countries (1,371 academic journal editors from CrossRef members, 1,263 academic journal editors from Nature Publishing Group, Oxford University Press, Cambridge University Press, and some scholarly societies, etc., and 671 recipients from the attendees of the 33rd Society for Scholarly Publishing Annual Meeting, 2011).
2. Survey Version 2 (SV2, part of the full

questionnaire), consisting of 10 questions (marked with * before numbers of questions in Appendix 1) chosen from SV1 because most of SV2 recipients without CrossRef membership would not have been able to respond to all of the SV1 questions, was sent to 607 non-native English speaking editors of academic journals covered by Web of Science, from Japan, South Korea, India, Singapore, China and Brazil etc. (Figure 1a).

We mostly sent the invitations to editors-in-chief and scientific editors although we had not asked the specific role of these journal editors in the questionnaire. (In other words, our questionnaire went to academics, subject experts who are directly involved in their respective journals and are likely to be responsible for selection of content. In Asia (e.g. China) the managing editor performs many of the roles of the scientific editor

our questionnaire went to academics, subject experts who are directly involved in their respective journals



* Life Sci. includes Bio-Sciences, Medicine and Agriculture, etc.; Chem./Phys./Eng. includes Mechanical/Civil/Environmental/Industrial/Control Engineering, Aerospace, Architecture, Mathematics and Statistics, etc.; Social Sci. includes Anthropology, Economics, Education, Geography, History, Law, Linguistics, Political Science, Public Administration, Psychology and Sociology, etc.; Computing/EE includes Computer Science, Electronics, Electrical Engineering (EE), Automation and Artificial Intelligence, etc.

Figure 1. All respondents analyzed by geographic journal location (a) and by discipline (b), showing CrossCheck users and non-users (Q1 and Q2 cross-analyzed with Q3, SV1 and SV2, $n = 219$).

in the West. The scientific editor/editor-in-chief is more of a figurehead, and so the questionnaire could have reached some managing editors.)

Data analysis

For multiple-choice questions, the percentage of each option was calculated and compared, and the results were cross-analyzed, as appropriate, by native language of the respondents, by discipline and by geographic journal location. The answers to Q6 and Q7 were analyzed using discriminant analysis (classification analysis) to classify original data, and finally to calculate their mean and standard deviation, respectively; for Q9, Q16, and Q21 the answers were calculated as the mean or median.

Results

1. Respondents by CrossCheck users, journal locations, and disciplines (SV1 and SV2, Q1–Q3)

The number of respondents to SV1 was 161, mainly from Western countries. Most of their journals are from some top publishers, such as Nature Publishing Group, IEEE, Elsevier, Springer, Wiley-Blackwell, Oxford University Press, Cambridge University Press, and some scholarly societies from the USA, which include some leading journals, such as *New England Journal of Medicine* (see full list in Appendix 2). And 95% of respondents show that their journals are published in English. There were 58 respondents to SV2; most of their journals are also among the leaders in their respective countries, and 93% of these are published in English. Overall, the response rate was 5.6 %, from 21 countries (Table 1 and Figure 1a).

The respondents in SV1 and SV2 are also

analyzed by different disciplines as shown in Figure 1b which shows that the replies from Life Sciences (Life Sci.) account for 43%, Chemistry/Physics/Engineering, etc. (Chem./Phys./Eng.) for 28%, Social Sciences (Social Sci.) for 15%, Computer Science/Electrical Engineering, etc. (Computing/EE) for 9%, and other disciplines for 5%.

2. Use of CrossCheck in checking the originality of submitted articles (SV1, Q4)

In SV1 and SV2, 42% of all respondents have experience of using CrossCheck: 51% of respondents to SV1, and 19% of respondents to SV2 (Table 1). In the field of Life Sciences, 49% respondents are CrossCheck users (Figure 1b). In addition, a few respondents indicated that they use a variety of other methods to detect possible plagiarism, such as eTBLAST, Medknow's plagiarism checking tool, Free Online Plagiarism Checker.

As Table 2 indicates, 32% of those SV1 respondents who use CrossCheck screen all submissions, while 34% screen only suspect papers, and 15% screen only accepted papers. The preference for screening only suspect papers is more marked in Chemistry/Physics/Engineering, etc., and Computer Sciences/Electrical Engineering, etc., than in Life Sciences and Social Sciences. Under the heading 'Other', some editors stated that CrossCheck could be used at any time if a paper aroused suspicion; others said that they use it to check only non-research articles that are almost without figures, equations and tables, or long papers. Other editors indicated that the screening was done by the publisher before submissions reached them; one mentioned that 'corresponding authors can choose to run their papers through CrossCheck (this is totally optional), and the journal pays the bill'.

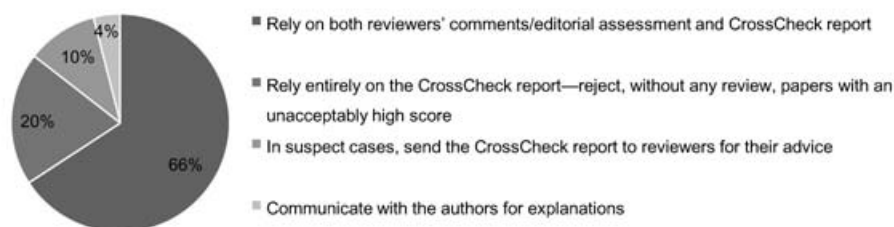
Table 1. Respondents to the survey (SV1 and SV2)

	No. of invitations	No. of respondents		Response rate (%)	% of CrossCheck users among respondents
		CrossCheck users	CrossCheck non-users		
Survey Version 1	3,305	82	79	4.9	51
Survey Version 2	607	11	47	9.6	19
Total	3,912	93	126	5.6	42

42% of all respondents have experience of using CrossCheck

Table 2. Which papers are screened using CrossRef (Q4, CrossRef users in SV1, n = 82)

Disciplines	Percentage of options (%)			
	All submissions	Only suspect papers	Only accepted papers	Other
Chem./Phys./Eng. (n = 26)	38	50	4	8
Life Sci. (n = 39)	31	28	18	23
Computing/EE (n = 5)	20	40	20	20
Social Sci. (n = 11)	27	18	27	27
Others (n = 1)	0	0	0	100
All (n = 82)	32	34	15	20

To what extent do journal editors rely on the CrossRef similarity report?**Figure 2. How respondents use the CrossRef similarity report (Q5, CrossRef users in SV1, n = 82).**

3. Use of CrossRef similarity reports (Q5–Q7)

Figure 2 indicates how respondents use the CrossRef similarity report. In SV1, 66% of respondents supplement it with expert opinions; 20% reject a paper out of hand if it is found to have unacceptably high similarity (Table 3 indicates what similarity index respondents consider to be ‘unacceptably high’); 10% forward the CrossRef report to the reviewers for their advice if it gives rise to suspicions of possible plagiarism, while 4% contact the authors to request an explanation.

Before giving the responses to Q6 and Q7, we need to define some aspects of the similarity report. The overall similarity index¹⁴ (OSI) is the ‘percentage of similarity between a submission and information existing in the iThenticate databases selected as search targets’. The single match similarity index¹⁴ (SMSI) is the percentage of similarity from a single source between a submission and information existing in the iThenticate databases selected as search targets. The OSI is one important indicator of a potentially plagiaristic paper; however, the degree of SMSI is also the other significant indicator. Table 3 shows the seriousness level of simi-

20% reject a paper out of hand if it is found to have unacceptably high similarity

Table 3. How respondents view the significance of the similarity index (Q6 and Q7, CrossRef users in SV1)

Seriousness (plagiarism/copying)	Suspected OSI, % (n=51)		Suspected SMSI, % (n=46)	
	Mean	Standard deviation	Mean	Standard deviation
Minor	8.99	4.23	8.99	4.23
Moderate	21.69	5.65	21.69	2.38
Serious	38.78	10.77	38.78	10.78
Trigger a reject	50.49	13.35	43.42	14.66
Trigger a rework	17.60	9.92	13.96	6.76

What are your views on verbatim or near-verbatim copying of a short extract from another work?

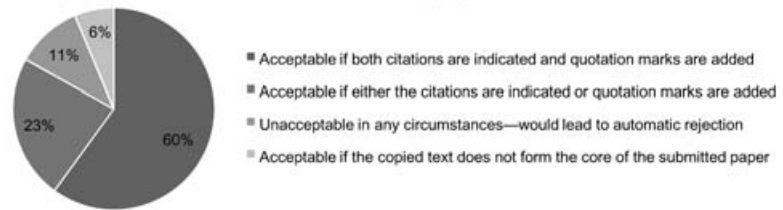


Figure 3. Respondents' views on verbatim, or near-verbatim, copying of a short extract from another work (Q8, SV1, $n = 160$).

larity that respondents feel suggests minor, moderate, or serious plagiarism. In the case of the OSI, the mode was $8.99 \pm 4.23\%$, $21.69 \pm 5.65\%$, and $38.78 \pm 10.77\%$, respectively; the mode index which would trigger outright rejection was $50.49 \pm 13.35\%$, while that for requesting revision by the author was $17.60 \pm 9.92\%$. In the case of the SMSI, the mode was $8.99 \pm 4.23\%$, $21.69 \pm 2.38\%$, and $38.78 \pm 10.78\%$, respectively (the results for the SMSI were much higher than expected, which may indicate that respondents were not clear what the SMSI is).

4. Attitudes to copying, cutting and pasting (Q8–Q14, Q19, and Q20)

4.1 Verbatim (or near-verbatim) copying (Q8–Q10)

Q8 deals with verbatim (or near-verbatim) copying of an extract from another work. As shown in Figure 3, 60% of respondents said that this can be acceptable provided it is clear that it is a quotation (e.g. quotation marks, indentation) and a full citation is given for the original source; 23% consider it acceptable with either quotation marks or a citation; 6% find it acceptable if the copied text does not form the core of the submitted paper, while 11% find it totally unacceptable and would always reject the paper.

Q9 deals with the length of word strings considered acceptable for verbatim copying, with or without citation. The data are cross-analyzed by discipline (Table 4). Respondents from Social Sciences show the lowest tolerance for word strings copied without citation. Respondents from the Life Sciences show considerably less latitude than other disciplines in the length of a quoted extract considered acceptable.

Table 4. Length of extract (number of words) considered acceptable for verbatim copying with and without citation (Q9, SV1, $n=138$)

Disciplines	No. of words (median) ^a	
	Without citation	With citation
Chem./Phys./Eng. ($n = 42$)	10	50
Life Sci. ($n = 54$)	8	30
Computing/EE ($n = 14$)	10	50
Social Sci. ($n = 28$)	1	50
All ($n = 138$)	7	50

^aData for the number of words are shown as a median as the responses varied widely.

Q10 asks about the policy regarding authors who cut-and-paste materials from other sources and integrate this with their own text. From Figure 4, it can be seen that 57% of respondents indicated that this would be unacceptable in all cases and the paper would be rejected, but 23% consider that cutting and pasting is acceptable if the paper is innovative, provided the author adds proper citations. Respondents from the Social Sciences are almost twice as likely to reject as those from the Life Sciences (Figure 4b). Chinese respondents are particularly disinclined to reject in these circumstances (Figure 4c).

4.2 Attitude and tolerance to copying in different sections of a paper (Q11–Q14)

Q11 and Q13 on editors' attitudes to the copied materials occurred in different parts of the articles. In Figure 5, the majority of respondents indicated that if between one-quarter and one-third of the content in the abstract, introduction or discussion is

respondents from Social Sciences show the lowest tolerance for word strings copied without citation

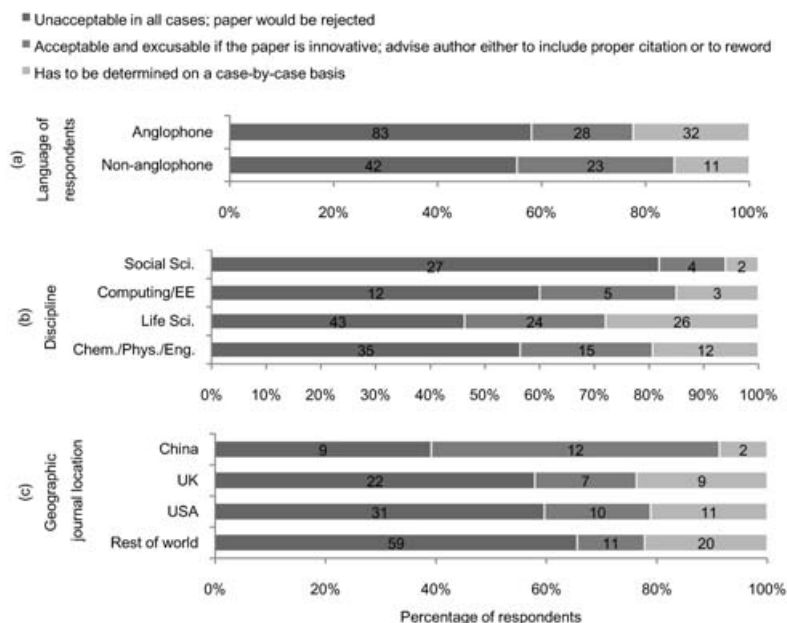


Figure 4. Policy of respondents regarding authors who cut-and-paste materials by language of respondents (a), by discipline (b), and by geographic journal location (c) (Q10, SV1 and SV2, $n = 219$).

In this figure and all subsequent bar chart the number in bars shows the actual numbers of respondents.

copied without citation, the paper is likely to be rejected. Respondents from the Life Sciences and Chemistry/Physics/Engineering, etc., are less likely to reject than the other respondents; they prefer to ask the author to include a proper citation, or to rewrite the content in his/her own words.

When cut-and-paste occurred in the materials and methods section of a paper (Q13), respondents generally indicated that this was unacceptable unless rewritten using the author's own words or with proper citation (Figure 6). Comments in response to this question suggested that decisions would

be based on journal requirements (e.g. 'some journals ask for a detailed description of the method'), article genre, type of text (e.g. 'some technical points are very difficult to re-word while retaining their meaning. It is also difficult to tell the difference between whether the person is copying from person X, or if both are copying from another source, like a text book'), length of the text (e.g. 'short string of purely technical detail about the methodology is OK'), or that advice would be sought from reviewers or the editor-in-chief.

Q12 and Q14 ask for editors' tolerance of

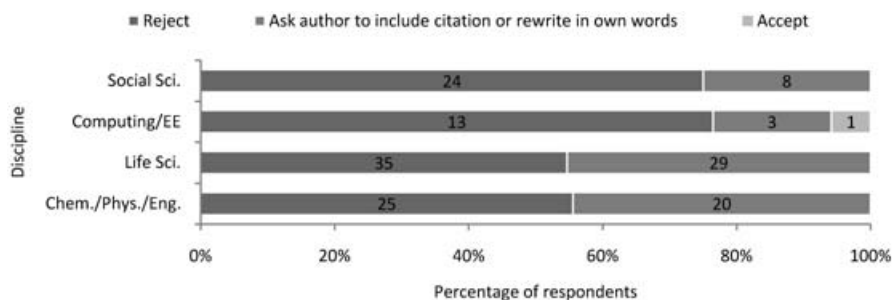


Figure 5. Attitude to copied content in abstract, introduction or discussion, by discipline (Q11, SV1, total response $n = 161$ which includes data of other disciplines ($n = 3$) not shown in this chart).

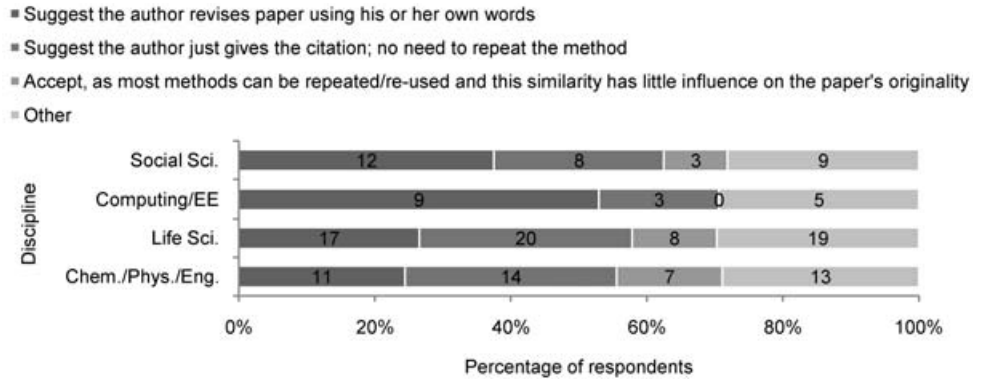


Figure 6. Attitude to copied content in materials and methods section without citation, by discipline (Q13, SV1, total response n = 161 which includes data of other disciplines (n = 3) not shown in this chart).

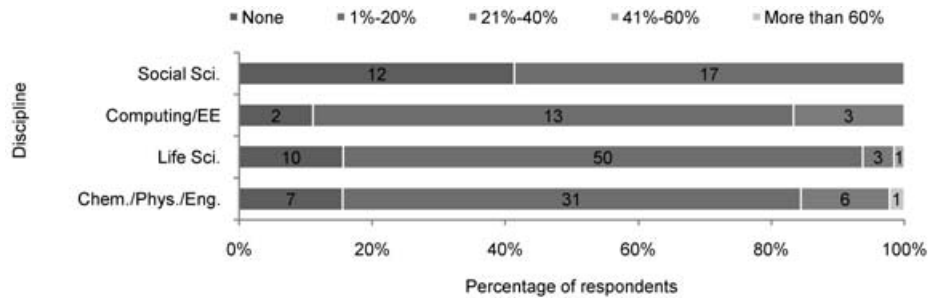


Figure 7. Percentage of copied content in the abstract, introduction or discussion considered acceptable with citation, by discipline (Q12, SV1, total response n = 161 which includes data of other disciplines (n = 5) not shown in this chart).

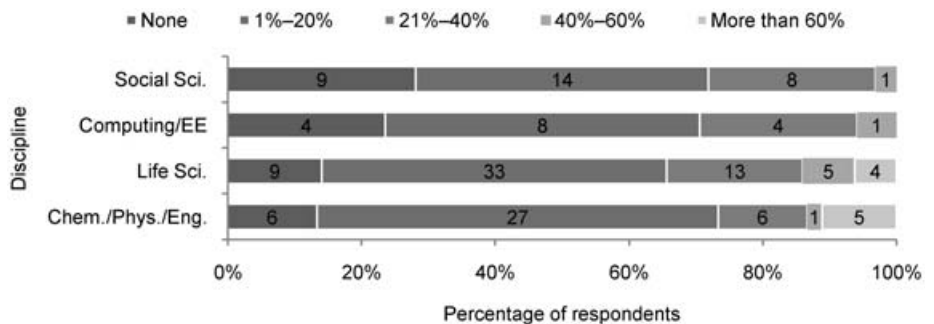


Figure 8. Percentage of copied content in the materials and methods section considered acceptable with citation, by discipline (Q14, SV1, total response n = 161 which includes data of other disciplines (n = 3) not shown in this chart).

90% of respondents suggest that even with citation the acceptable percentage of copied content is very low, i.e. 1–20%

copied content in the abstract/introduction/discussion and the materials and methods section, respectively. In the case of the abstract, introduction and discussion, more than 90% of respondents suggest that even with citation the acceptable percentage of copied content is very low, i.e. 1–20%

(Figure 7). In the case of the materials and methods section, about 70% of respondents indicated that copied content should be 20% or below. However, nearly 20% (13/64) from Life Sciences would tolerate the copied contents of 21–40% (Figure 8).

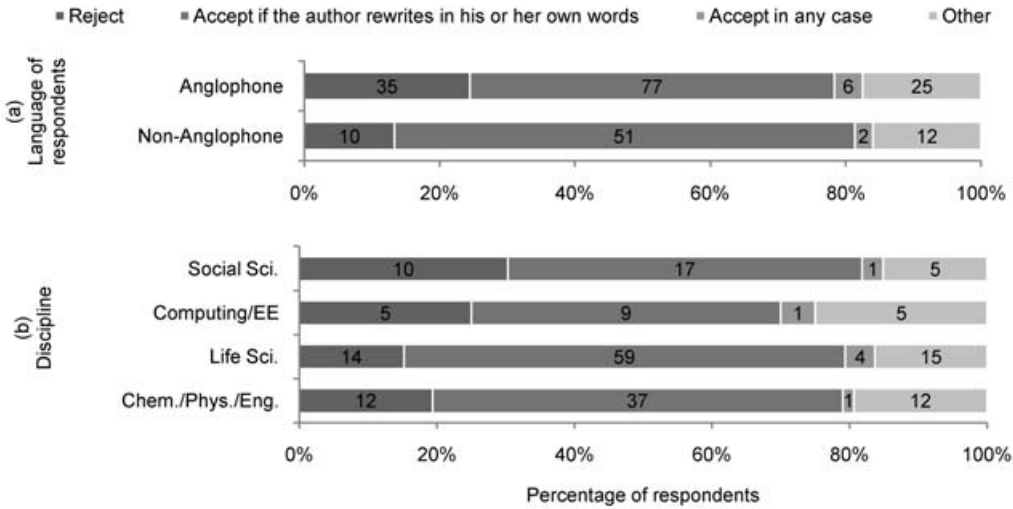


Figure 9. Respondents' course of action when the author of a review article has summarized previously published papers wholly or mainly in the original authors' own words, by language of respondents (a) and by discipline (b) (Q19, SV1 and SV2, $n = 218$).

4.3 Review articles consisting predominantly of copied text 'cutting-and-pasting' (Q19 and Q20)

Summaries of the papers discussed, often using the original authors' own words, are unsurprisingly more common in review articles; Q19 and Q20 address the respondents' views about the acceptability of cutting-and-pasting in review articles. From the respondents in different disciplines and languages, we can see an average 79% of them responding to Q19 suggested that even review articles would be rejected in their current form, or accepted only after rewriting in the review author's own words, if the summaries consisted wholly or mainly of the

original authors' words. A total of 25% of Anglophone respondents would reject such papers out of hand; for non-Anglophone respondents this percentage was 13% (Figure 9a). There were no marked disciplinary differences (Figure 9b).

78% of respondents to Q20 felt that review papers with an OSI $> \sim 50\%$ would not be acceptable, and more than 60% of respondents said that the acceptable OSI in a review article would be $< 35\%$. Again, disciplinary differences were not marked (Figure 10). 16% choose 'others' showing no metrics, which depends on what is similar and why.

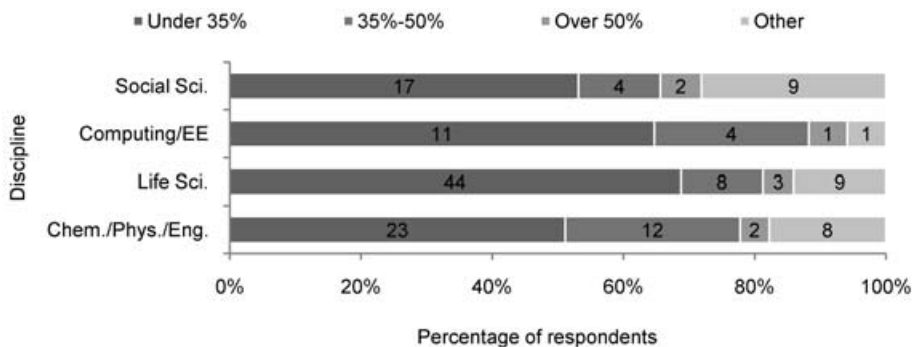


Figure 10. Percentage figure for the overall similarity index considered acceptable in a review article, by discipline (Q20, SV1, total response $n = 161$ which includes data of other disciplines ($n=3$) not shown in this chart).

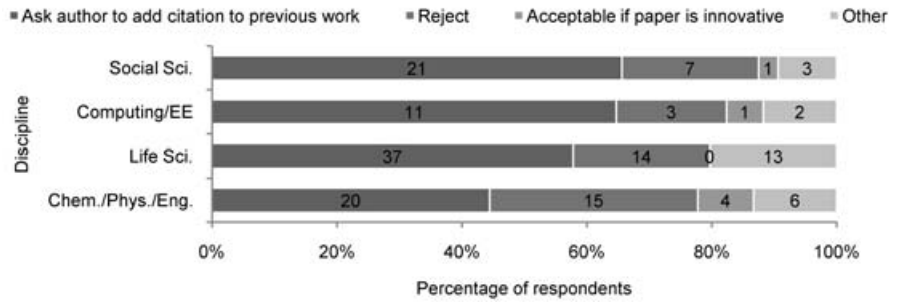


Figure 11. Respondents' view of self- or team plagiarism in the results and conclusions sections without citation, by discipline (Q15, SV1, total response n = 161 which includes data of other disciplines (n = 3) not shown in this chart).

far fewer respondents in Computer Science/Electrical Engineering were inclined to reject out of hand than those in other disciplines

5. Self-/team plagiarism and duplicate publication (Q15–Q18)

5.1 Self-plagiarism and team plagiarism

Is it ethical or reasonable that ‘author or co-authors reuse their own previously written work or data in a “new” written product without letting the reader know that this material has appeared elsewhere’?¹⁵ Q15, Q17, and Q18 deal with self-plagiarism and team plagiarism, which means copying from each other a great deal within the same research program or group.⁸ While the definition of self-plagiarism is difficult, and although self-plagiarism does not involve the theft of someone else’s work, it is still likely to contravene journal policy, and often also violates the original publisher’s copyright.¹⁶

Plagiarism by an individual of his/her previous work (self-plagiarism) and plagiarism of the work of the team in which the author is a member are very similar; in both cases the author is copying his/her own or his/her group’s previously published tables, figures and text with no or few changes, without making clear what has been copied and without citing the previous publication. The responses to Q15, which refers to self- or team plagiarism in the results and conclusions sections of a paper, indicate that the most common response is to ask the author to provide a citation to his/her previous work; there are few differences between disciplines (Figure 11).

Q17 asked ‘How do you deal with an article whose title, aims and methodology

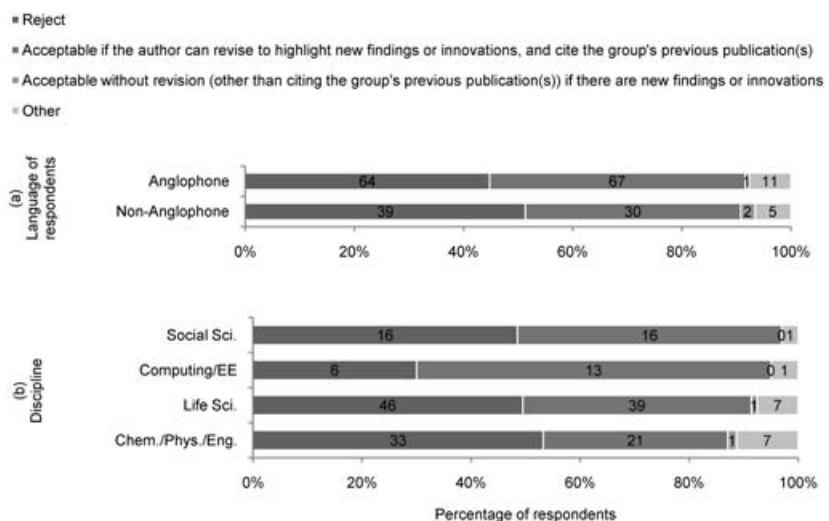


Figure 12. Respondents' view of suspected team plagiarism, by language of respondents (a) and by discipline (b) (SV1 and SV2, Q17, n = 219).

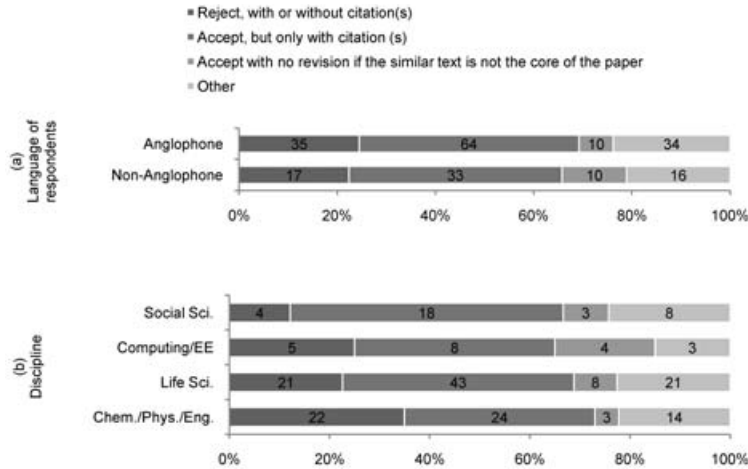


Figure 13. Respondents' course of action in cases of significant self-plagiarism, where the author(s) claim that the papers are a series of studies, by language of respondents (a) and by discipline (b) (Q18, SV1 and SV2, $n = 219$).

are identical or highly similar to those of another paper published by the same research group (team), and where only the specific examples and materials, etc. are different?' 91% of editors from both Anglophone and non-Anglophone countries responded that they would either reject such a paper out of hand, or accept it only if revised to highlight new findings or innovations, and citing the group's previous publication(s) (Figure 12a). Far fewer respondents in Computer Science/Electrical Engineering were inclined to reject out of hand than those in other disciplines (Figure

12b). Some answers mentioned other factors, such as the referee's opinion, the author's explanation, and the amount of significant additional content.

Q18 asks how the respondent would react if the author(s) claim that the papers are a series of studies with the same background, which will inevitably lead to similarity in the text. 44% of respondents said they would accept but only with citation, while 24% said they would reject with or without citation. Irrespective of language and discipline (Figure 13), there is a strong consensus (total of 68%) either to reject, or to accept only with

'Should papers previously published in conference proceedings legitimately be republished in journals?'

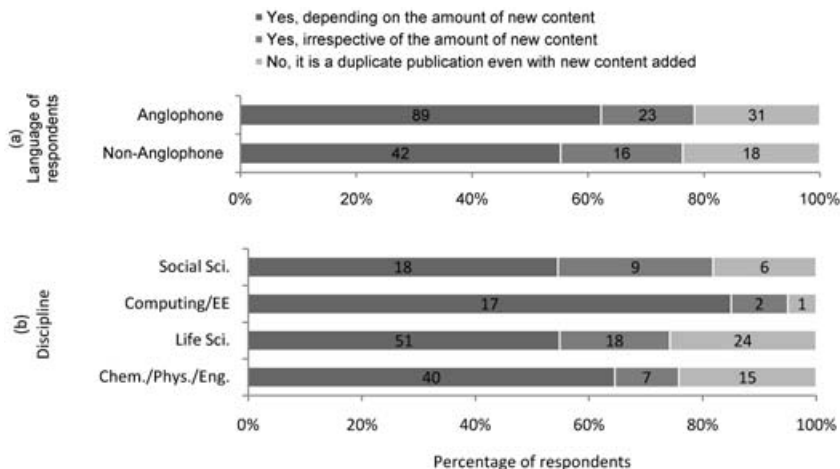


Figure 14. Respondents' view on whether conference proceedings papers can legitimately be republished in a journal, by language of respondents (a) and by disciplines (b) (Q16, SV1 and SV2, $n = 219$).

Table 5. Amount of new content considered necessary by respondents in order to justify republication of papers from conference proceedings (Q16, SV1 and SV2, n = 131)

Disciplines	n	% of new content (mean)
Chem./Phys./Eng.	40	50
Life Sci.	51	45
Computing/EE	17	35
Social Sci.	18	45
Others	5	43
All	131	46

proper citation. A number of respondents commented that the author must rephrase or revise with proper citation, minimizing repetition, and eliminating all cutting and pasting to avoid self-plagiarism. One respondent mentioned that the reviewer's or editor-in-chief's comments on the innovation or originality of the paper should also be taken into account; others stressed that if the similarity is in a core area of the article (such as in the results section) with no substantially new overall contribution, the article should be rejected out of hand.

5.2 Republication of papers from conference proceedings (Q16, SV1 and SV2)

Q16 addresses the question 'Should papers previously published in conference proceedings legitimately be republished in journals?' 60% of respondents think such papers can properly be republished provided they include new content (Figure 14a,b). And they indicated that there should be no less than 46% of new material (Table 5). However, 22% of respondents considered this to be duplicate publication, even with new content added.

The lowest percentage opting for rejection out of hand came from the field of Computer Science/Electrical Engineering, etc. (5%), and more than 80% of editors in this field indicated that a proceedings paper can be republished if it includes new content.

6. Percentage of papers rejected because of plagiarism (Q21)

Q21 asks respondents to give the approxi-

Table 6. Percentage of papers submitted to respondents' own journals that are rejected specifically on account of plagiarism (Q21, SV1 and SV2, n = 153)

Respondents by language	n	% rejected specifically on account of plagiarism (mean)
Anglophone	102	3.5
Non-Anglophone	51	11.0
All	153	5.9

Table 7. Respondents' willingness to state in their 'Instructions for Authors' that they use CrossCheck to scan all papers submitted (Q22, SV1, n = 148)

Option	%
Already state that we use CrossCheck to scan for plagiarism	45
Would consider stating that we use CrossCheck	28
Do not (or not yet) state that we use CrossCheck	8
Would not state that we use CrossCheck: seems intimidating and counterproductive	19

mate percentage of papers rejected because of plagiarism. From the answers to Anglophone (n = 102) and non-Anglophone (n = 51) editors, we find the rejection rates depending on plagiarism based on each journal's experience are about 3.5% and 11.0%, respectively (Table 6).

7. Willingness to refer to use of CrossCheck in 'Instructions to Authors' (SV1, Q22)

Q22 asks whether respondents would be willing to refer to their use of CrossCheck in the 'instructions to authors'. As shown in Table 7, 45% of respondents (SV1, n = 148) indicated that they already did so; a further 28% would consider doing so. However, 19% thought that it would be intimidating and counterproductive, and 8% would not (or not yet) consider such a statement.

Respondents' approach to suspicion of plagiarism varies relatively little between journals from developed and developing countries, and between journals from Anglophone and non-Anglophone countries,

the survey data suggest a strong consensus among all respondents about the criteria for determining plagiarism

even though the percentage of articles rejected on the grounds of plagiarism by journals from non-Anglophone countries is almost double that of journals from Anglophone countries.

Discussion

1. What is and is not acceptable

‘There are tools to detect non-originality in articles, but instilling ethical norms remains essential.’¹⁷ The *Merriam-Webster Dictionary* defines plagiarize as ‘to commit literary theft: present as new and original an idea or product derived from an existing source’.³ A number of guides to research publication^{4,18,19} make it clear that taking text (even a phrase or a sentence) from someone else’s work without acknowledgement is theft of their intellectual property, and hence constitutes plagiarism, whether it is deliberate or not. The survey data suggest a strong consensus among all respondents about the criteria for determining plagiarism; this was more marked than the local variations (whether based on language, discipline, or geographical journal location).

Respondents to the survey generally agree that verbatim (or near-verbatim) copying must include a citation and direct quotation marks, and that no more than seven words should be copied without citation and quotation marks, and the strictest response is in social sciences, only one word (Table 4). Respondents in Chemistry/Physics/Engineering, etc., and in Computer Science/Electrical Engineering and Social Sciences are of the opinion that even if a citation and quotation marks are given, no more than 50 words should be copied; respondents in the Life Sciences felt that citations of up to 30 words were acceptable.

In a review article, the author aims to digest a wide range of previous papers and synthesize her/his findings to form a coherent argument about a topic or a focused description of a field. 78% of respondents felt that review papers with an OSI > ~50% would not be acceptable, and 60% felt that those with an OSI > 35% would not be acceptable (Figure 10).

Even when repeating ‘common knowl-

edge’^{20,21} in some sections of a paper, such as classical methods in biomedical procedures, respondents felt that authors should wherever possible use their own words. Only one respondent felt that between one-third and one-quarter cut-and-pasted content could be acceptable in the abstract, introduction or discussion sections (Figure 5), but 11% felt that this was acceptable in the material and methods sections (Figure 6).

There is a great temptation to plagiarize one’s own work or that of one’s team, since the number of publications is often used as an indication of a researcher’s scientific merit.²² However, team plagiarism has long been criticized by journal editors and publishers.²³⁻²⁵ In this study, in the opinion of all groups, whether by discipline, language, or geographic location, over 90% of editors were highly consistent in their disapproval of team plagiarism.

As to whether self-plagiarism can be defined as a type of plagiarism, there are a lot of discussions in the iThenticate White Paper.¹⁶ And in our survey, it is encouraging that the majority of editors would either reject a substantially self-plagiarizing article (24%), or accept it but only with the addition of proper citations (44%). To the question ‘is it possible to steal from oneself’, Hexam pointed out ‘the essence of self-plagiarism is the author’s attempts to deceive the reader’.²⁶

Republication of papers which have previously appeared in conference proceedings is a difficult issue; there are circumstances in which republication of papers (with the permission of the copyright owner) has been considered acceptable in the past. However, there is now a huge amount of information available via the Internet. Having similar or duplicate content can cause confusion and waste publishing resources. For example, in academic journal publishing, as far as we know most journals are published online with digital object identifiers (DOIs). A republished paper with the same content could have two DOIs that will result in misleading information and waste the reader’s time. 60% of respondents to the survey felt that proceedings papers could properly be republished in journals provided they

republishing of papers which have previously appeared in conference proceedings is a difficult issue

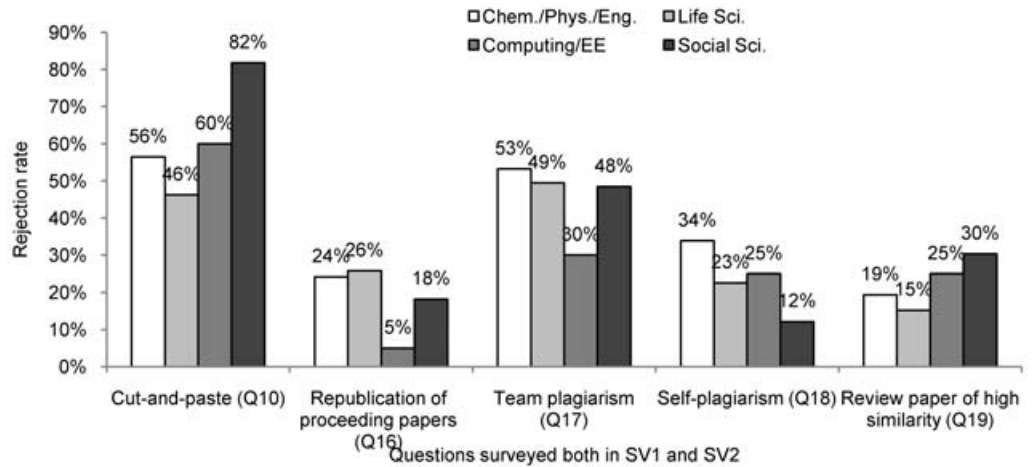


Figure 15. Differences in those choosing 'rejection' in response to five typical questions in different disciplines.

global editors have expressed a strong mainstream view about ethical standards

contained at least 46% of new content (Figure 14, Table 5).

2. Differences by discipline and language

2.1 Disciplinary differences in rejection decisions

For five typical plagiarism problems such as cut-and-paste (Q10), duplication of conference proceedings (Q16), self-plagiarism (Q17), team plagiarism (Q18), review papers of high similarity (Q19), based on the rejection decisions' percentage of respondents from these subjects, we can find differences between different disciplines (Figure 15).

We find that the Social Sciences show the lowest tolerance, with 82% and 30% rejection rates, respectively to cut-and-paste (Q10) and much copied material in review papers (Q19), but to self-plagiarism (Q18) they show a high tolerance, with a rejection rate of only 12%. What is the reason? This may be related to the characteristics of social science articles, with greater expression of a personal viewpoint in the text. Perhaps because the the writing process in the social sciences is itself a creative one, social science editors express the lowest tolerance here. Further exploration of self-plagiarism may be needed. As the iThenticate White Paper on the ethics of self-plagiarism states: 'writers maybe unaware of the ethics and laws involved in reusing or repurposing their

own texts',¹⁶ although our journal editors do show they recognize the problem by using the plagiarism detection tool.

The other noticeable phenomenon is that Computer Science/Electrical Engineering shows the lowest rejection rate in republication of proceedings papers (Q16), and team plagiarism (Q17), with 5% and 30%, respectively. Because this subject mainly depends on the updating of new technologies and team co-operation, there are more conference proceedings publications.

The attitude toward the five questions in Chemistry/Physics/Engineering and Life Sciences disciplines seems to be very similar.

2.2 Language differences

(a) Majority consensus

Figure 16a illustrates the extent of consensus between respondents from English speaking countries (mainly in SV1, $n = 143$) and non-native English speaking countries (mainly in SV2, $n = 76$), to a number of key questions. Especially for team plagiarism, both groups show over 91% 'reject' rate or 'acceptable if can revise to highlight new findings or innovations, and cite the group's previous publication(s)'. And the percentages of the other four mainstream questions also are also over 55%, which proves global editors have expressed a strong mainstream view about ethical standards.

some authors from developing countries may be unaware that they are committing plagiarism and infringing copyright law

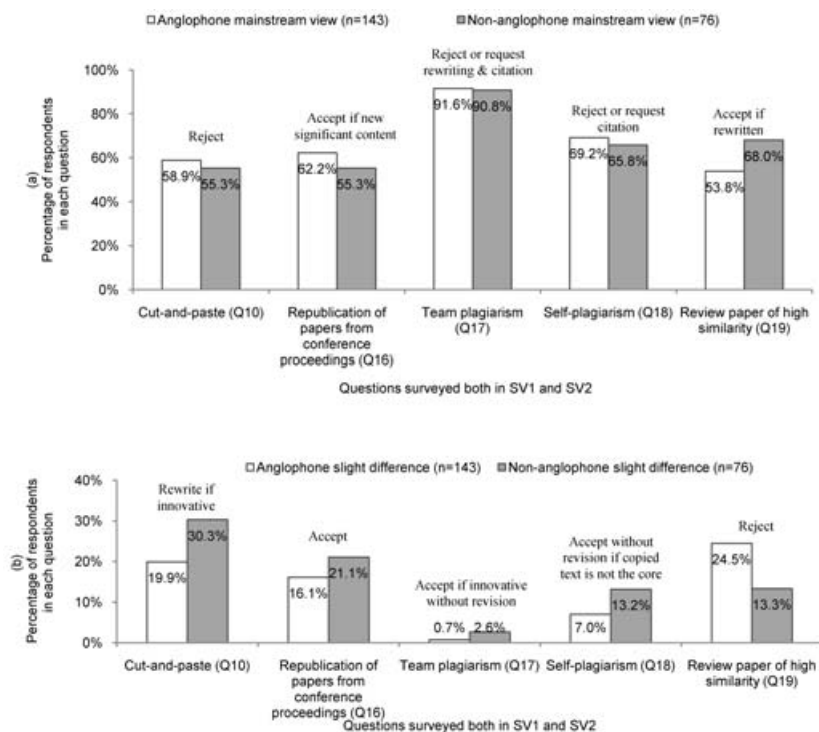


Figure 16. Mainstream view (a) and small differences (b) to five problems between Anglophone ($n = 143$) respondents and non-Anglophone ($n = 76$) respondents.

(b) Difference in minority opinions

However from Figure 16b, we know there are small variations that show the attitude of non-Anglophones is a little less rigorous than that of the Anglophones. These differences may be due to cultural and language differences arising from the wide range of social perspectives and stages of national development²⁷. Copyright law has been well-documented for more than 300 years in Western countries,^{28,29} whereas it has been established much more recently in developing countries (e.g. in China the international copyright law has been in effect only since 1991). Thus, some authors from developing countries may be unaware that they are committing plagiarism and infringing copyright law.²⁹ It may take some time before authors from developing countries catch up with Western countries, and before the incidences of plagiarism can effectively be reduced in those countries.

Conclusions

The main results of this survey can be sum-

marized as follows: (1) The plagiarism detection tool and similarity report are very useful and effective, and can assist editors to screen documents suspected of plagiarism. (2) Global editors have expressed a strong mainstream view of ethical standards even though there are slight variations between different disciplines and countries, as well as between non-Anglophone editors and Anglophone editors. (3) Given those variations, perhaps a global principle and practical approaches to prevent plagiarism and duplicate publication should be established.

Any attempt to reduce or prevent plagiarism and duplicate publication will require that not only academic journal editors or scholarly publishers, but also scientists themselves, whether acting as authors or as reviewers, accept the responsibility to raise their own standards, and indeed to establish criteria so that the next generation can clearly understand the difference between ethical and unethical publishing behavior.²⁷

In addition, our survey indicates a relatively high awareness of plagiarism issues,

our survey indicates a relatively high awareness of plagiarism issues

and high ethical standards in dealing with them, among academic journal editors around the world. However, the survey was limited in that it only approached relatively few, leading, journals and reported on a small sub-set of global journals. In some ways undertaking a survey itself can be seen as a form of advocacy for higher ethical standards; it may be worthwhile undertaking a more comprehensive survey.

Appendices

The survey questionnaire breakdown of respondents are available as an online supplement to this paper.

Acknowledgements

We would like to express our sincere thanks to all respondents in our survey. We especially thank Sally Morris, Ian McIntosh, and referees for their professional guidance. We would also like to thank PhD student Jiangsheng Yi in the Department of Mathematics, Zhejiang University for his statistical analysis. Finally, we appreciate our editorial team, Hanfeng Lin, Ziyang Zhai, and Xinxin Zhang for their active participation and discussion of the work.

References

1. 'Plagiarism, n.' 2012. *Oxford English Dictionary Online*, 3rd edn. Available at <http://www.oed.com>.
2. 'Plagiarize, v.' 2012. *Oxford English Dictionary Online*, 3rd edn. Available at <http://www.oed.com>.
3. 'Plagiarize.' 2012. *Merriam-Webster Dictionary*. Available at <http://www.merriam-webster.com/dictionary/plagiarize>.
4. Hames, I. *Peer Review and Manuscript Management in Scientific Journals: Guidelines for Good Practice*, 1st edn. Oxford, Wiley-Blackwell, 2007, pp.177–181.
5. Butler, D. 2010. Journals step up plagiarism policing cut-and-paste culture tackled by CrossRef software. *Nature*, 466: 167. <http://dx.doi.org/10.1038/466167a>
6. *Nature Comments*, 2012. How to stop plagiarism. *Nature*, 481: 21–23. <http://dx.doi.org/10.1038/481021a>
7. Meddings, K. Credit where credit's due: plagiarism screening in scholarly publishing. *Learned Publishing*, 23: 5–8.
8. Zhang, Y.H. 2010. CrossRef: an effective tool for detecting plagiarism. *Learned Publishing*, 23: 9–14. <http://dx.doi.org/10.1087/20100103>
9. Zhang, Y.H. 2010. Chinese journal finds 31% of submissions plagiarized. *Nature*, 467: 153. <http://dx.doi.org/10.1038/467153d>
10. Griffin, C. 2010. The Journal of Bone & Joint Surgery's CrossRef experience. *Learned Publishing*, 23: 132–135. <http://dx.doi.org/10.1087/20100208>
11. Kleinert, S., on behalf of the editors of all Lancet journals, 2011. Checking for plagiarism, duplicate publication, and text recycling. *The Lancet*, 377: 281–282. [http://dx.doi.org/10.1016/S0140-6736\(11\)60075-5](http://dx.doi.org/10.1016/S0140-6736(11)60075-5)
12. Available at <http://www.crossref.org/crosscheck/members.html>
13. COPE Research Granted, 2011. CrossRef Guidance: An Analysis of Typical Cases of Plagiarism in Different Disciplines. Available at <http://publicationethics.org/resources/research>
14. Available at <http://www.crossref.org/crosscheck/CrossCheckManual.pdf>.
15. Roig, M. 2006. Avoiding plagiarism, self-plagiarism, and other questionable writing practices: a guide to ethical writing. Available at <http://facpub.stjohns.edu/~roigm/plagiarism/Self%20plagiarism.html>.
16. iThenticate White Paper, 2011. The Ethics of Self-Plagiarism.
17. *Nature* Editorial, 2010. Plagiarism pinioned. *Nature*, 466: 159–160. <http://dx.doi.org/10.1038/466159b>
18. Greetham, B. *How to Write Better Essays*, 2nd edn. Basingstoke, Palgrave Macmillan, 2008.
19. McMillan, K. and Weyers, J. *How to Write Essays & Assignments*, 2nd edn. Pearson Education Limited, 2011.
20. Purdue University Online Writing Lab, 2010. Is It Plagiarism Yet? Available at <http://owl.english.purdue.edu/owl/resource/589/02/> (accessed 10 October 2011)
21. Cal State San Marcos Library, 2008. How to Avoid Plagiarism? Available at http://library.csusm.edu/plagiarism/howtoavoid/how_avoid_common.htm (accessed 10 October 2011)
22. Franzoni, C., Scellato, G., and Stephan, P. 2011. Changing incentives to publish. *Science*, 333: 702–703. <http://dx.doi.org/10.1126/science.1197286>
23. Benos, D.J., et al., 2005. Ethics and scientific publication. *Advances in Physiology Education*, 29: 59–74. <http://dx.doi.org/10.1152/advan.00056.2004>
24. Lin, H.F., Jia, X.Y., Zhang, Y.H., Zhang, C.J., Jin, M.Q., Zhang, X.X., and Zhai, Z.Z., 2011. Guarantee academic originality: Duty of journal editors – workflow and analysis with CrossRef of JZUS (A/B/C). *Chinese Journal of Scientific and Technical Periodicals*, 22: 328–333 (in Chinese).
25. Radulescu, G. 1985. Duplicate publication is boring. *American Journal of Diseases of Children*, 139: 119–120.
26. Hexham, I. 1999. The Plague of Plagiarism. Department of Religious Studies. The University of Calgary. Available at <http://c.faculty.umkc.edu/cowande/plague.htm#self>.
27. Suresh, S. Moving toward global science. *Science*, 333: 802. <http://dx.doi.org/10.1126/science.1210025>
28. Williams, A. 2007. Copyright: past, present and future. *Learned Publishing*, 20: 113–119. <http://dx.doi.org/10.1087/174148507X185072>
29. Thatcher, S.G. 2008. China's copyright dilemma. *Learned Publishing*, 21: 278–284. <http://dx.doi.org/10.1087/095315108X356707>

Yuehong (Helen) ZHANG*

Xiaoyan JIA

Journal of Zhejiang University-SCIENCE

(A/B/C)

Zheda Road 38, Hangzhou 310027

PR China

*Email: jzus@zju.edu.cn

Learned Publishing, 25
doi:10.1087/20120412

Appendix 1 – Survey questionnaire

Title: Survey on Detecting Plagiarism in Journals using CrossRef

This survey is being carried out as part of a research project funded by the Committee on Publication Ethics (COPE).

(SV1 contains 22 questions, of which 10 (marked with *) were used in SV2)

***Q1 What is the subject of your journal(s)? Choose firstly a broad category, as below.**

- Chemistry/Physics/Engineering (Mechanical/Civil/Environmental/Industrial/Control, Aerospace etc.), Architecture, Mathematics/Statistics
- Life Sciences (including Bio-Sciences, Medicine, Agriculture)
- Computer Science/Electronics/Electrical Engineering/Automation/Artificial Intelligence etc.
- Social Sciences (Anthropology/Economics/Education/Geography/History/Law/Linguistics/Political Science/Public Administration/Psychology/Sociology)
- Others

***Q2 Basic information about your journal(s)**

Country
Language

***Q3 Do you use CrossRef?**

- Yes
- No

Q4 How do you use CrossRef in checking the originality of submitted articles?

- All submissions are CrossChecked
- Only accepted papers are CrossChecked
- Only suspect papers are CrossChecked
- Other (please specify)

Q5 As a journal editor, to what extent do you rely on the CrossRef similarity report to judge whether submitted papers involve plagiarism?

- Rely entirely on the CrossRef report—reject, without any review, papers with an unacceptably high score
- Rely entirely on reviewers' comments; do not consider CrossRef report
- Rely on both reviewers' comments and CrossRef report
- In suspect cases, send the CrossRef report to reviewers for their advice
- Other (please specify)

***Q6 The overall similarity index percentage is one important indicator of a potentially plagiaristic paper.**

Please indicate at what percentage you decide the paper contains:

- Minor plagiarism, minimum %
- Moderate plagiarism, minimum %
- Serious plagiarism, minimum %
- Triggers a reject, minimum %
- Triggers request to author to rework it, minimum %

(for SV2, similar Q is 'In a journal paper, what percentage of copied content would you consider acceptable with citation? ...')

Q7 The degree of similarity for each single match is also significant. Please indicate for single matches at what percentage you decide the paper contains:

- Minor plagiarism, minimum %
- Moderate plagiarism, minimum %
- Serious plagiarism, minimum %
- Triggers a reject, minimum %
- Triggers request to author to rework it, minimum %

Q8 What are your views on verbatim or near-verbatim copying of a short extract from another work?

- Acceptable if the copied text does not form the core of the submitted paper
- Acceptable if both citations are indicated and quotation marks are added
- Acceptable if either the citations are indicated or quotation marks are added
- Unacceptable in any circumstances—would lead to automatic rejection

Q9 What length of extract (number of words) would you consider acceptable for verbatim copying in the following two cases?

- Without citation (number of words)
- With citation (number of words)

***Q10 What is your policy regarding authors who cut-and-paste materials from other sources and integrate it with their own text?**

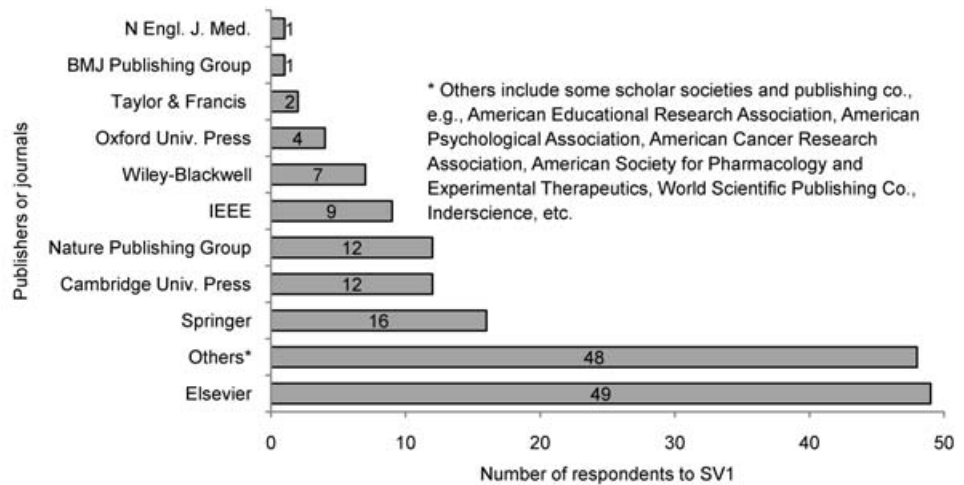
- Acceptable and excusable if the paper is innovative; advise author either to include proper citation or to rewrite in own words
- Unacceptable in all cases; paper would be rejected
- Other (please specify)

Q11 In sections Abstract/Introduction/Discussion, if between 1/4 and 1/3 of the content is copied without citations, what would you do?

- Reject
- Ask author to include citation or rewrite in own words
- Accept

Q12 In sections Abstract/Introduction/Discussion, what percentage of copied content would you consider acceptable with citation?

- None
- 1%–20%
- 21%–40%
- 41%–60%
- More than 60%



Appendix 2. List of respondents to SV1, by publisher ($n = 161$).

Q13 In section Materials & Methods, if between 1/4 and 1/3 of the content is copied without citations, what would you do?

- Accept, as most methods can be repeated/re-used and this similarity has little influence on the paper's originality
- Suggest the author revises paper using his or her own words
- Suggest the author just gives the citation; no need to repeat the method
- Other (please specify)

Q14 In section Materials & Methods, what percentage of copied content would you consider acceptable with citation?

- None
- 1%–20%
- 21%–40%
- 41%–60%
- More than 60%

Q15 In section Results and Conclusions, what is your view of authors copying their own previously published tables or figures with no or small changes without citation?

- Reject
- Ask author to add citation to previous work
- Acceptable if paper is innovative
- Other (please specify)

***Q16** Do you think papers previously published in conference proceedings can legitimately be republished in a journal with the addition of new content?

- No, it is a duplicate publication even with new content added
- Yes, irrespective of the amount of new content
- Yes, depending on the amount of new content. Please indicate what amount of new content as a minimum percentage

***Q17** How do you deal with an article whose title, aims and methodologies are identical or highly similar to those of another paper published by the same research group, and where only the specific examples and materials, etc. are different?

- Reject
- Acceptable if the author can revise to highlight new findings or innovations, and cite the group's previous publication(s)
- Acceptable without revision (other than citing the group's previous publication(s)) if there are new findings or innovations
- Other (please specify)

***Q18** Authors sometimes reuse significant portions of their own work, either verbatim or near-verbatim (self-plagiarism); they may claim that the papers are a series of studies with the same background, which will inevitably lead to similarity in the text. What is your reaction in cases like this?

- Reject, with or without citation(s)
- Accept, but only with citation (s)
- Accept with no revision if the similar text is not the core of the paper
- Other (please specify)

***Q19** In writing a review paper, authors necessarily summarize (and cite) previously published papers. How do you handle cases where they have predominantly used the original authors' own words?

- Reject
- Accept in any case
- Accept if the author rewrites in his or her own words
- Other (please specify)

Q20 In a review paper, what percentage figure for the overall similarity index would you accept?

- Under 35%
- 35%–50%

- Over 50%
- Other (please specify)

*Q21 In your own journal(s) and hence subject area, approximately what percentage of papers you receive are rejected specifically on account of plagiarism? (We are not asking for the title of your journal(s) to avoid any possible malevolent use of this information)

Q22 In your “Instructions for Authors”, would you consider announcing that you use CrossCheck to scan all papers submitted, or would this seem intimidating and counterproductive?

Appendix 2 – see figure: List of respondents to SV1, by publisher (n = 161)

(Note: The number of respondents to SV2 was a little small and most of their journals are published independently. Therefore, the publishers of these journals in SV2 were not classified.)