



* As part of a research project funded by the Committee on Publication Ethics (COPE)

A Global Survey on Detecting Plagiarism in Journals using CrossCheck/iThenticate

Helen (Y.H.) ZHANG, Managing Editor of JZUS (A/B/C)
Speaker in 2011 CrossRef Annual Member Meeting

The survey results have been published in *Learned Publishing*, 25: 292–307, 2012
Y.H(Helen) Zhang & X.Y. Jia "A survey on the use of CrossCheck for detecting plagiarism in journal articles "

<http://alpsp.publisher.ingentaconnect.com/content/alpsp/lp/2012/00000025/00000004/art00008>



www.zju.edu.cn/jzus

jzus@zju.edu.cn



Aim: we hope to learn in the survey

- 1. How do journal publishers/editors worldwide use CrossCheck/iThenticate and analyze the similarity index?**
- 2. What are journal publishers/editors' attitude & tolerance toward typical plagiarism in different disciplines?**
- 3. What are mainstream views and differences between editors in western countries and non-western countries?**

Survey Version 1 (SV1) contains 22 questions, of which 10 were used in Survey Version 2 (SV2, marked with *) because most of SV2 recipients without Crosscheck members would not have been able to respond to all of the SV1 questions.



Q1 ~ Q3. Respondents to the survey by CrossCheck users, Countries, Disciplines, Languages

	Invitation	Number of respondents		Response rate	Percentage of CrossCheck users among respondents
		CrossCheck user	CrossCheck non-user		
SV1	3305*	82	79	4.9%	51%
SV2	607**	11	47	9.6%	19%
Total	3912	93	126	5.6%	43%

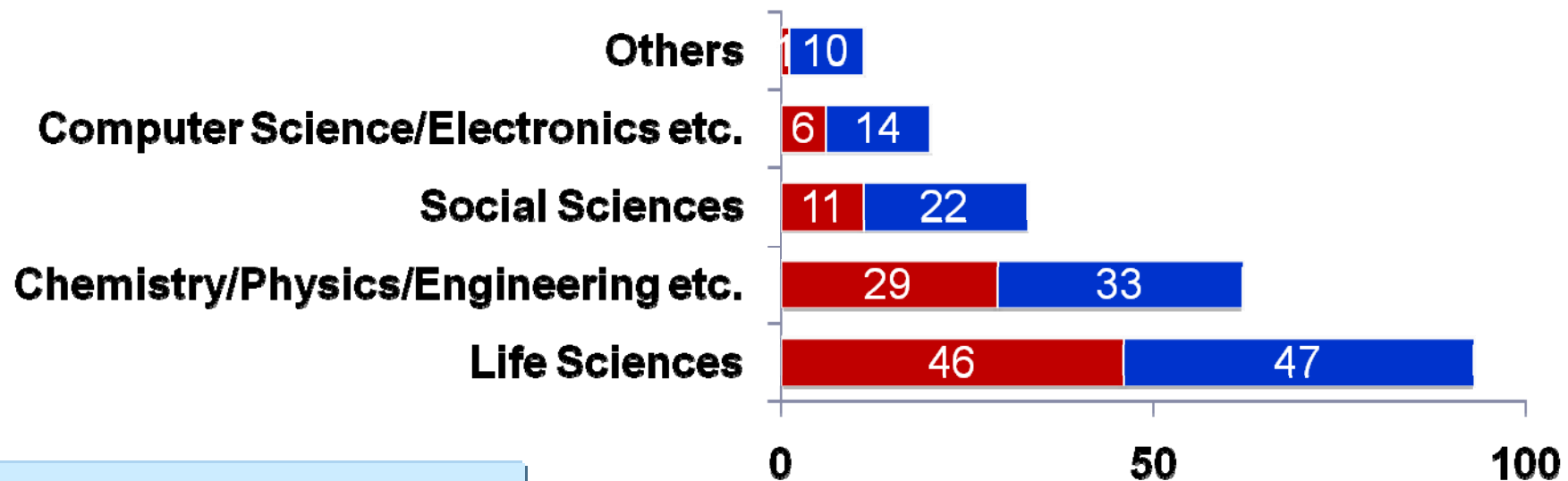
SV1: 3305* invitations: mostly from native English-speaking countries in the Western world, including CrossCheck journal members (1371) + leading journal editors (1263) + attendees of 33rd Society for Scholarly Publishing Annual Meeting (671). **95% of replies are English journals**

SV2: 607** invitations are from non-native English speaking countries' editors of Journals covered by Web of Science (**Leading journals in their countries**), from Japan, South Korea, India, Singapore, China, and Brazil etc. **93% of these are published in English.**



Q1. By discipline, CrossCheck users and non-users (cross-analyzed, n=219)

■ CrossCheck user ■ CrossCheck non-user



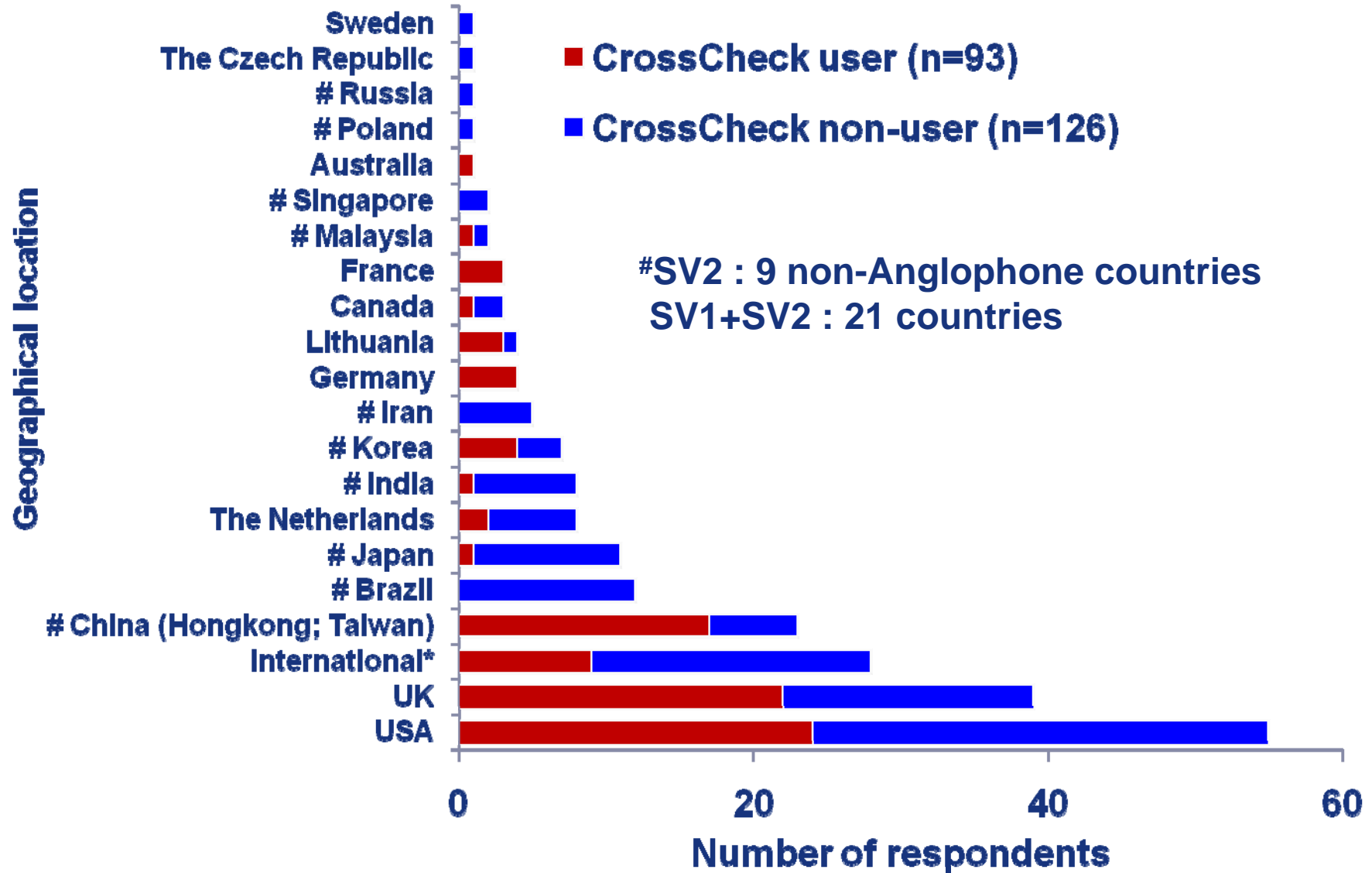
User %, by discipline

Life sci.	50% (46/93);
Chem/Phy/Eng.	47% (29/62);
Soc sci.	33% (11/33)
Computer/EE.	30% (6/20);
Other	9% (1/11)

Number of respondents

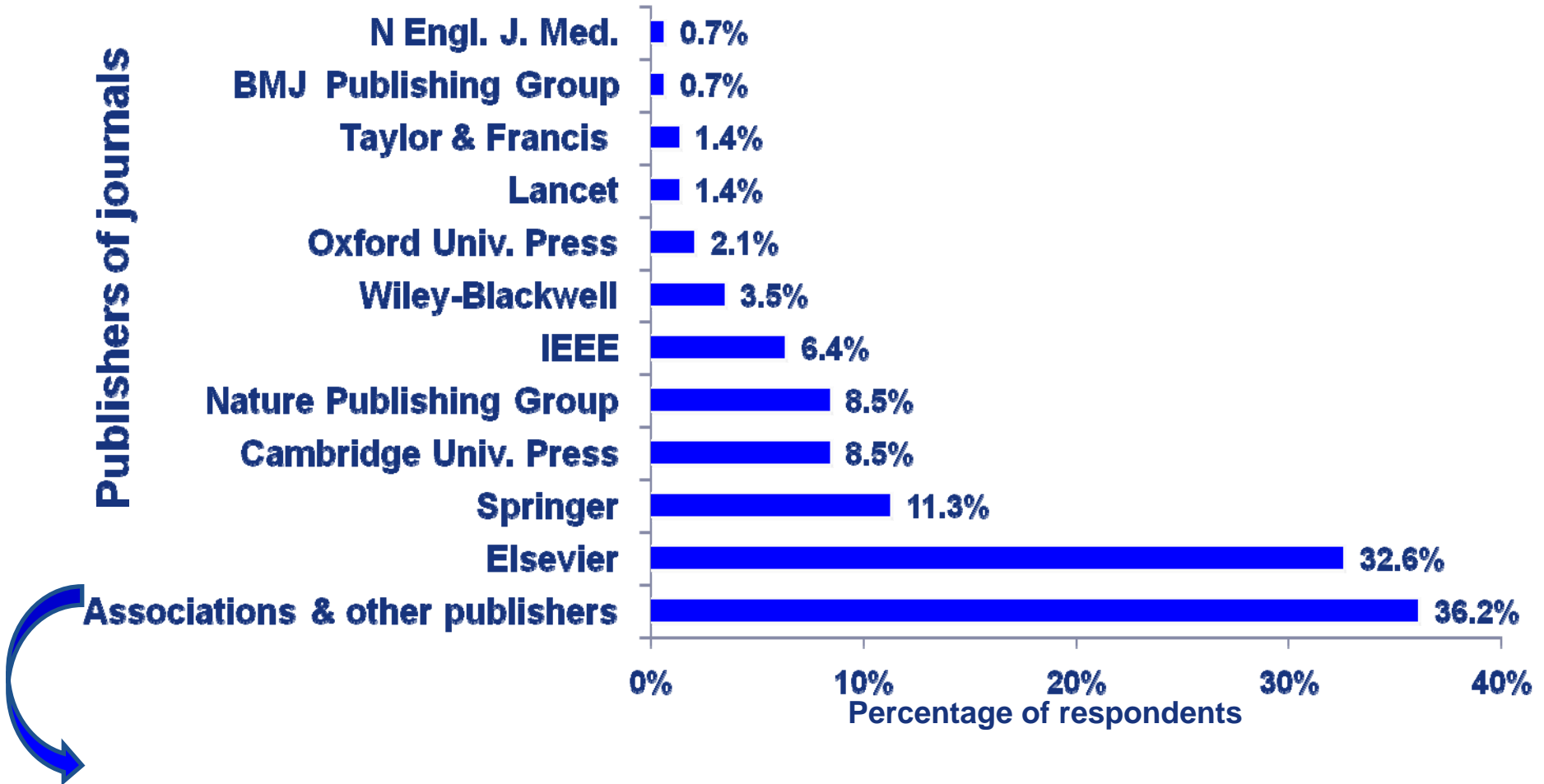


Q2 & Q3. By geographical location (SV1 & SV2), CrossCheck users and non-users (cross-analyzed, $n=219$)





Publishers of respondents' journals in SV1 ($n=161$)



Association & other publishers: American Educational Research Association, American Psychological Association, American Cancer Research Association, Future medicine, World Scientific Publishing Co., Inderscience, ASPET, HighWire Press, and BioOne, etc.



Q4. How do you use CrossCheck in checking the originality of submitted articles (SV1)?

Table 2. Which papers are screened using CrossCheck (*n=82 users to SV1*)

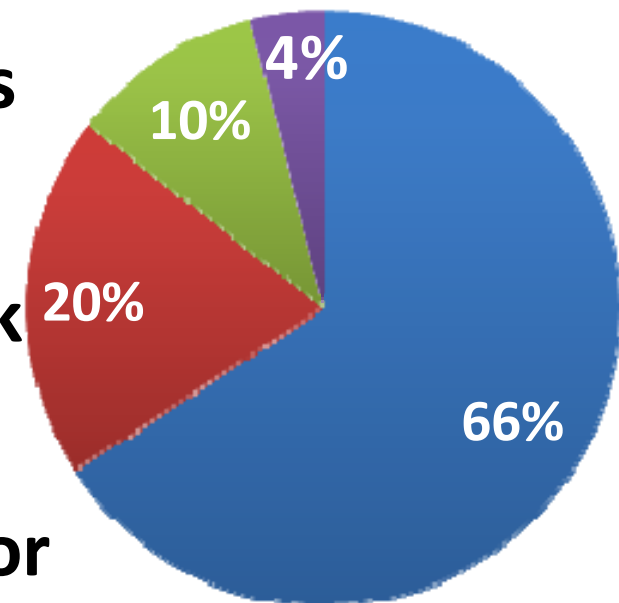
Respondent, by discipline	Percentage of options			
	All submissions	Only suspect papers	Only accepted papers	Other*
Chem./Phys./Eng., etc. (<i>n=26</i>)	38%	50%	4%	8%
Life Sci. (<i>n=39</i>)	31%	28%	18%	23%
Computer Sci., etc. (<i>n=5</i>)	20%	40%	20%	20%
Social Sci. (<i>n=11</i>)	27%	18%	27%	27%
All (<i>n=82</i>)	32%	34%	15%	20%

*Other: At any time if a paper aroused suspicion; check only non-research; the screening was done by the publisher before submissions reached them; one mentioned that “corresponding authors can choose to run their papers through CrossCheck, we pay the bill



Q5. To what extent do you rely on the CrossCheck similarity report to judge whether submitted papers involve plagiarism (SV1)?

- **Rely on both reviewer's comments & CrossCheck**
- **Rely on CrossCheck: reject papers with unacceptable high score**
- **In suspect cases, send CrossCheck report to reviewers for advice**
- **Communicate with the authors for explanations**



CrossCheck users responding to SV1, *n*=82



Q6 & Q7. How do you categorize the OSI and SMSI scores in terms of seriousness? (SV1)

What is Overall Similarity Index (OSI) ?

The total percentage of similarity between a submission and information existing in the CrossCheck/iThenticate databases selected as search targets.

What is Single Match Similarity Index (SMSI)?

The percentage of similarity from a single source between the iThenticate database and the submitted document.

The overall similarity index (OSI, Q6) is one important indicator of a potentially plagiaristic paper; However, the degree of single match similarity index (SMSI, Q7) is also significant indicator.

Important warning: The plagiarism screening tool & similarity report are extremely useful and effective, but they are not replacements for editorial and review expertise.



Q6 & Q7. How do you categorize the OSI and SMSI scores in terms of seriousness (SV1)?

Results of CrossCheck users responding to SV1

Seriousness (Plagiarism/Copying)	Minimum OSI (n=51)		Minimum 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%
Triggers a rejection	50.49%	13.35%	43.42%	14.66%
Triggers a reworking	17.6%	9.92%	13.96%	6.76%

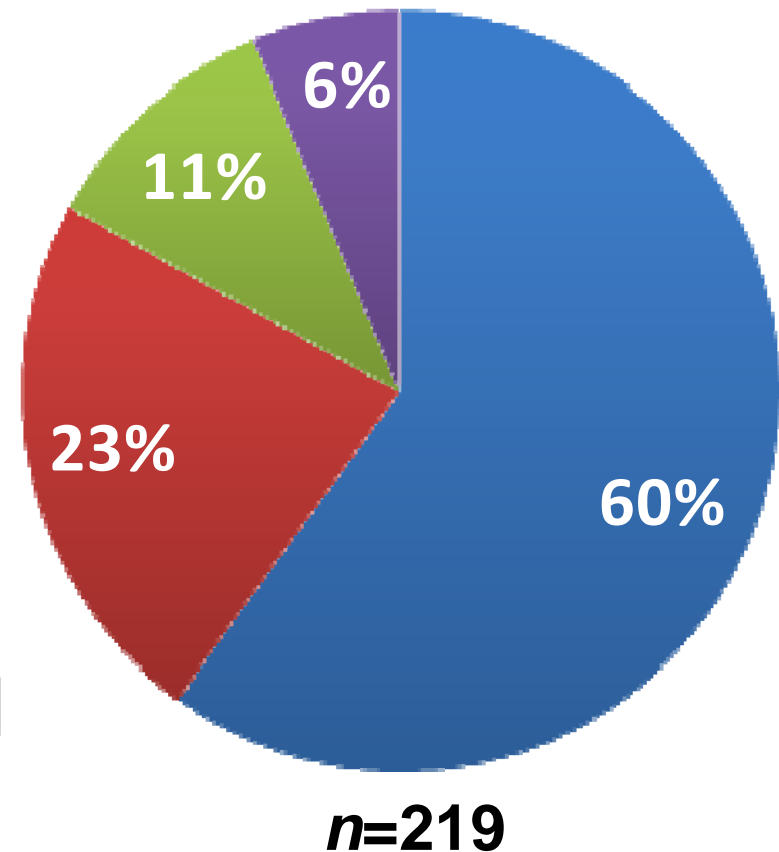
*We use the discriminant (classification) analysis to statistic original data.

The results of the SMSI are much higher than we expected, which maybe indicates a lack of understanding of what the SMSI is.



Q8. What are your views about verbatim or near-verbatim copying from another work? (SV1 & SV2)

- **Acceptable if both citation and quotation**
- **Acceptable if either citation or quotation**
- **Unacceptable in any circumstances-reject**
- **Acceptable if the copied text is not the core**





Q9. What length of extract (number of words) is acceptable for verbatim copying with or without citation (SV1)?

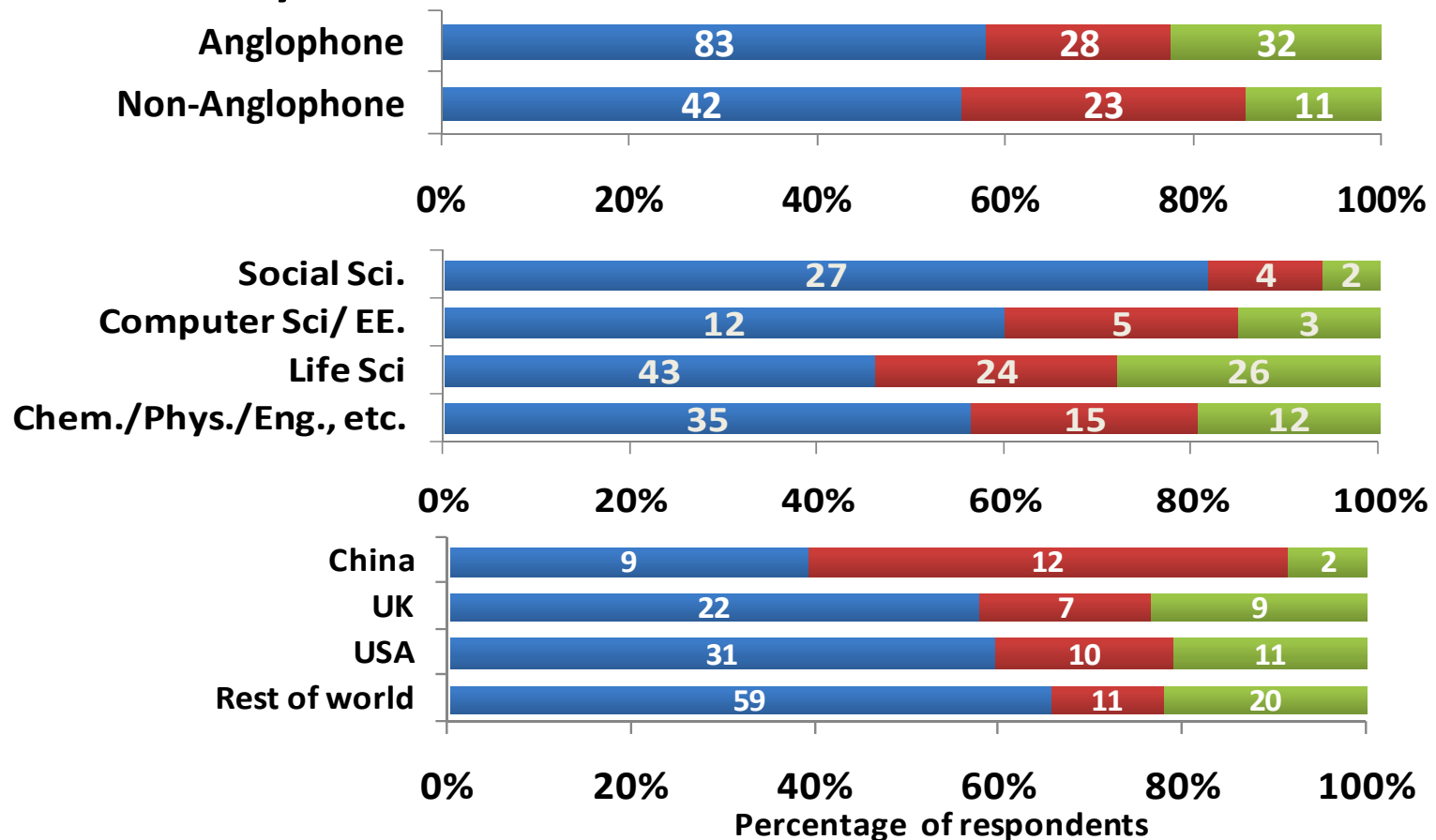
Respondent, by disciplines	Number of words (median)*	
	Without citation	With citation
Chem./Phys./Eng., etc. (<i>n</i> =42)	<10	<50
Life Sci. (<i>n</i> =54)	<8	<30
Computer Sci., etc. (<i>n</i> =14)	<10	<50
Social Sci. (<i>n</i> =28)	<1	<50
All (<i>n</i>=138)	<7	<50

* Median shown as responses covered a wide range



Q10. What policy regarding authors who cut-and-paste materials from other sources and integrate it with their own text (SV1 & SV2)?

- Unacceptable in all cases—reject
- Acceptable if innovative—add citation or reword
- Determined case-by-case





Q11 & Q12. Where copying occurs in the paper: abstract, introduction or discussion (SV1)

Q11 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

Attitude

Q12 What percentage of copied content would you consider **acceptable with citation**?

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

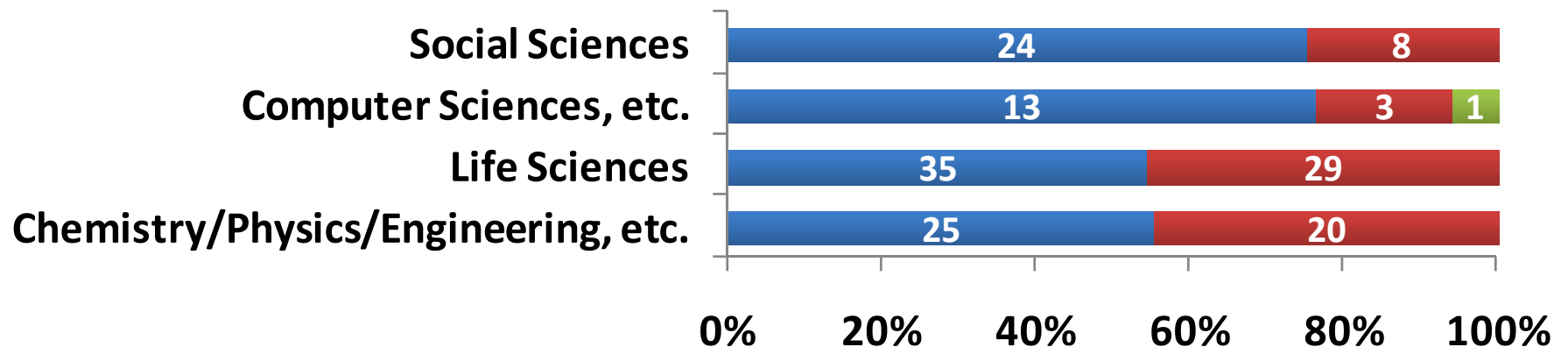
Tolerance



Q11 Attitude to copied content in Abstract, Introduction or Discussion, by discipline (SV1, n=161)

Q11 Attitude

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



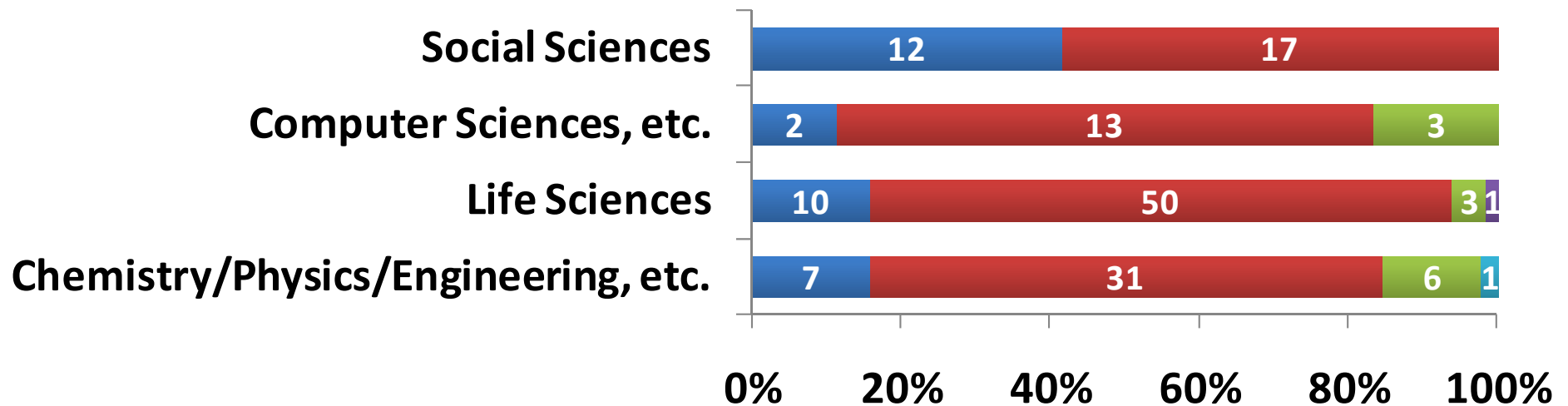
Only one respondent (0.6%) felt that this was acceptable in the Abstract, Introduction or Discussion sections



Q12 Tolerance of copied content in Abstract, Introduction or Discussion, by discipline (SV1, n=161)

Q12 Tolerance

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

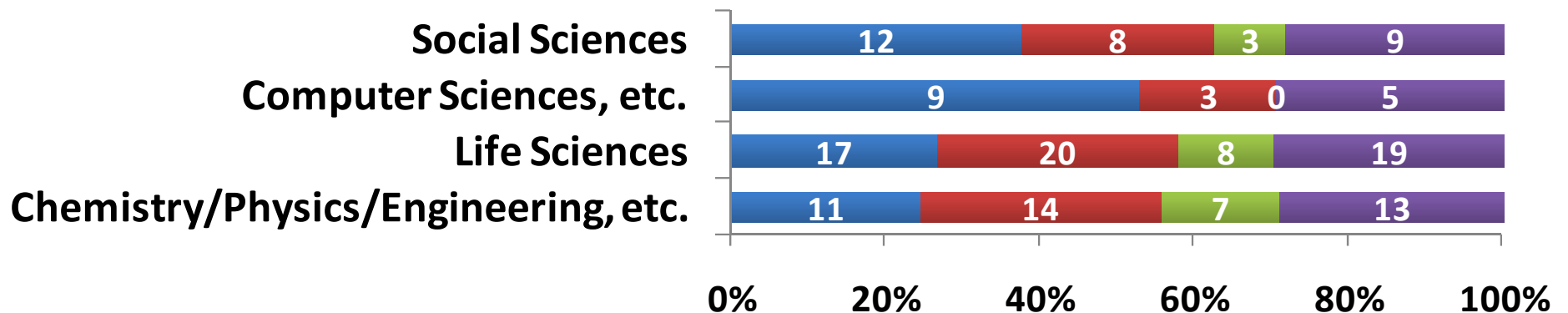


About 90% suggest that even with citation the acceptable percentage of copied content should be less than 1%~20%



Q13 Attitude to copied content in Materials & Methods (SV1, $n=161$)

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

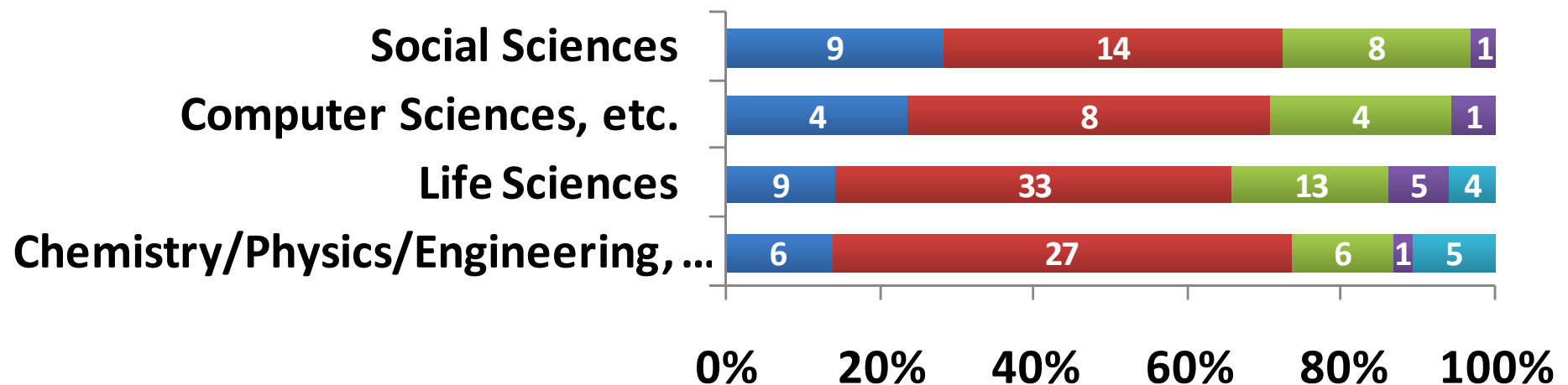


89% indicated: unacceptable unless rewritten using the author's own words or with proper citation



Q14. Tolerance of copied content in Materials & Methods (SV1, $n=161$)

■ None ■ 1%–20% ■ 21%–40% ■ 40%–60% ■ More than 60%

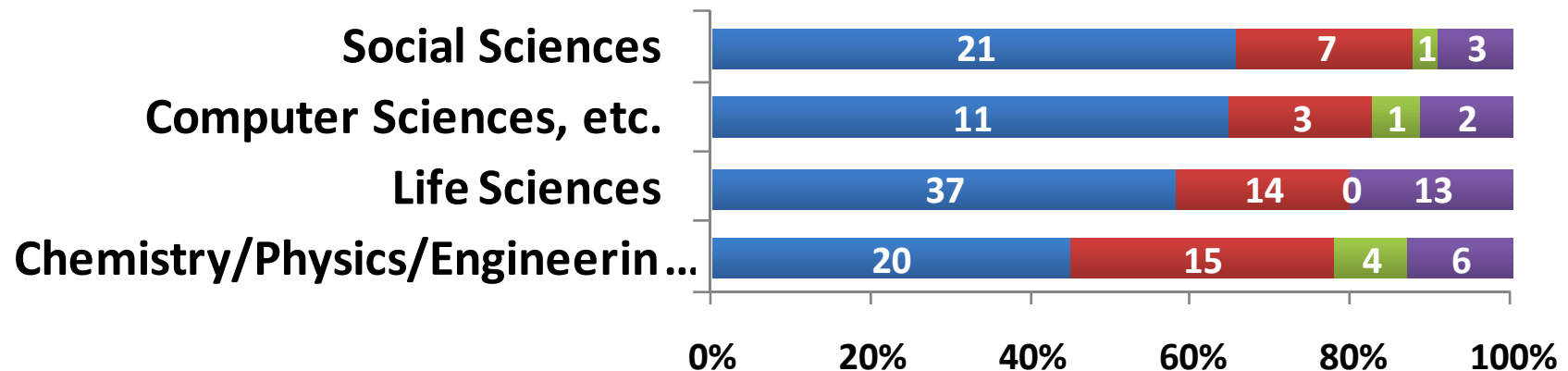


70% of respondents indicated that copied content should be 20% or below. In social sciences this tolerance was very low, probably because materials and methods sections are rare. However near 20% from life and computer sciences can tolerate the copied contents to be high 21%~40% in this section.



Q15. What attitude to authors reusing their own published Tables or Figures Without Citation? (SV1, n=161)

- Ask author to add citation to previous work
- Reject
- Acceptable if paper is innovative
- Other

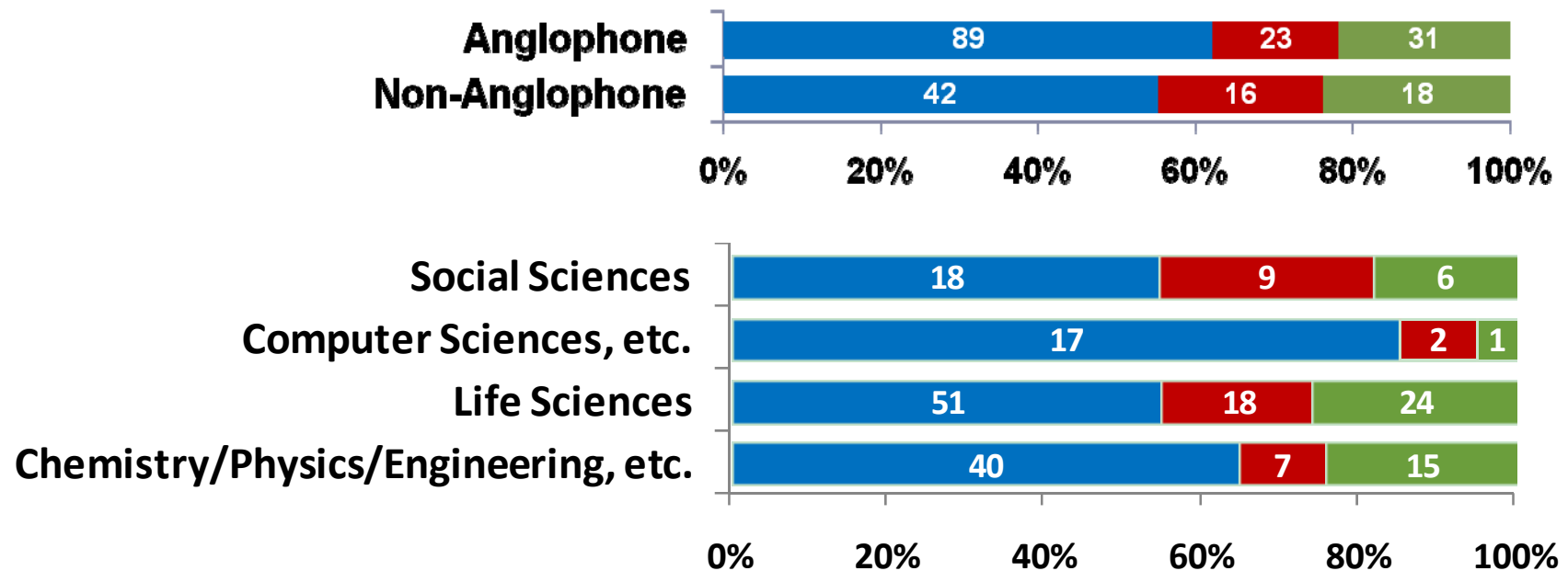


79% of respondents' attitudes to Q15 is concerned with asking the author to give a citation for his/her previous work or reject. There is little difference between disciplines.



Q16. Attitudes to “republishing” of papers from conference proceedings legitimately? (SV1 & SV2, n=219)

- Yes, depending on the amount of new content
- Yes, irrespective of the amount of new content
- No, it is a duplicate publication even with new content added



60% of respondents think such papers can properly be republished provided they include new content. And they indicated that there should be no less than 45% of new material. However, 22% of respondents who considered this to be duplicate publication, even with new content added.



Q16. Tolerance: How much new content here? At lowest 46 % ? (SV1 & SV2, n=219)

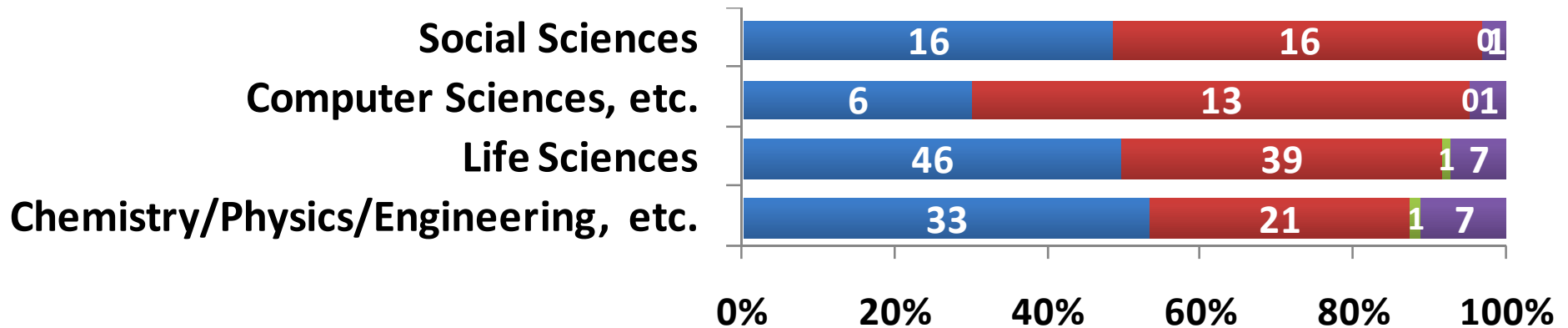
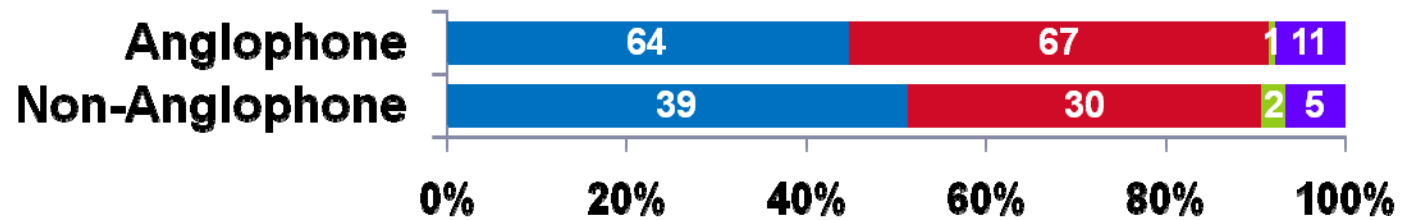
Amount of new content considered necessary by respondents in order to justify **republishing** of papers from **conference proceedings**

Respondents, by disciplines	<i>n</i>	New content percentage (mean, %)
Chemistry/Phys./Eng., etc.	40	> 50
Life Sci.	51	> 45
Computer Sci./EE	17	> 35
Social Sci.	18	> 45
All	131	> 46



Q17. Team plagiarism: response to articles with same/similar title, aims & methodologies to others from same group (SV1 & SV2, n=219)

- Reject**
- Acceptable if innovative & citing group's previous works**
- Acceptable without revision if innovative**
- Other**

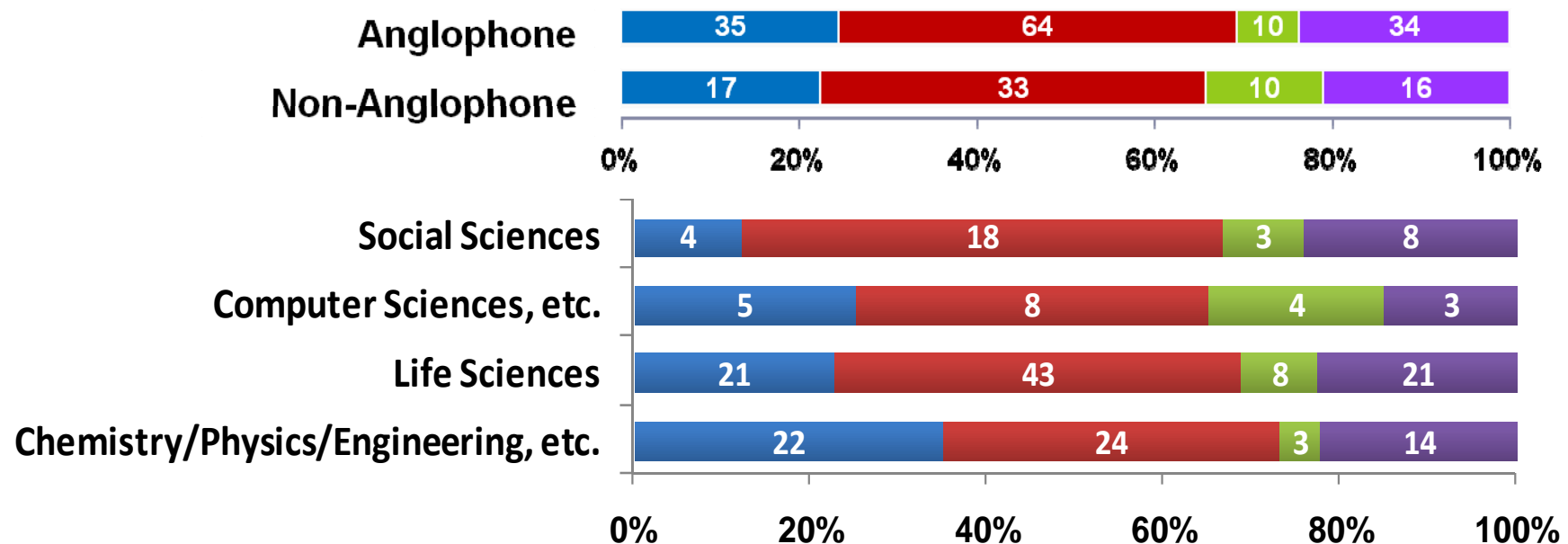


Team plagiarism has long been criticized by editors. And in this study the views from disciplines and languages, over 90% editors were consistent to against team plagiarism.



Q18. Self-plagiarism: response to claim that similar papers are part of a series of studies (SV1 & SV2, $n=219$)

- **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**

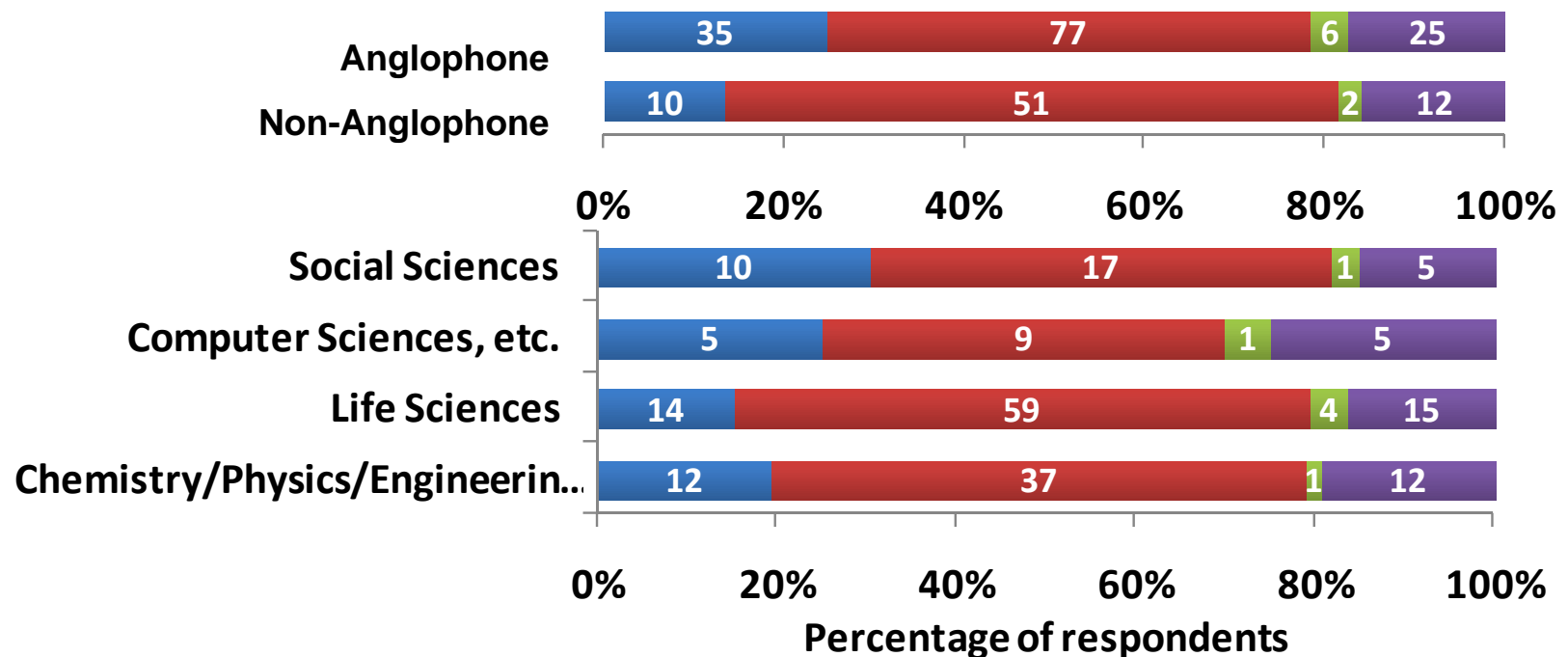


For self-plagiarism, even it is still a controversial definition, but in our survey, it is encouraging that nearly 70% of editors have recognized its seriousness and choose “reject”, or “accept but only with proper citations”.



Q19. Review articles consisting predominantly of copied text “cutting-and-pasting ” (SV1 & SV2, $n=219$)

- **Reject**
- **Accept if the author rewrites in his or her own words**
- **Accept in any case**
- **Other**

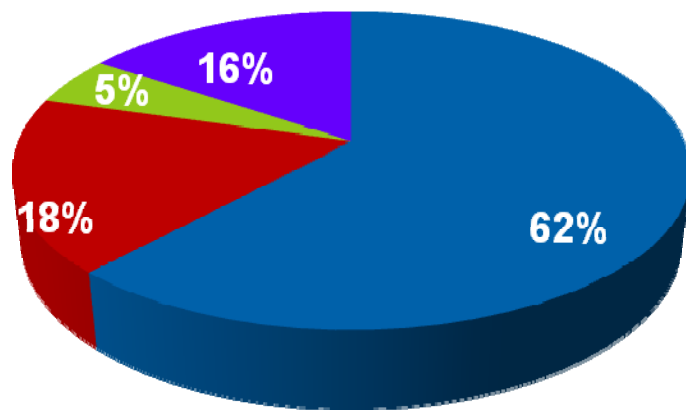
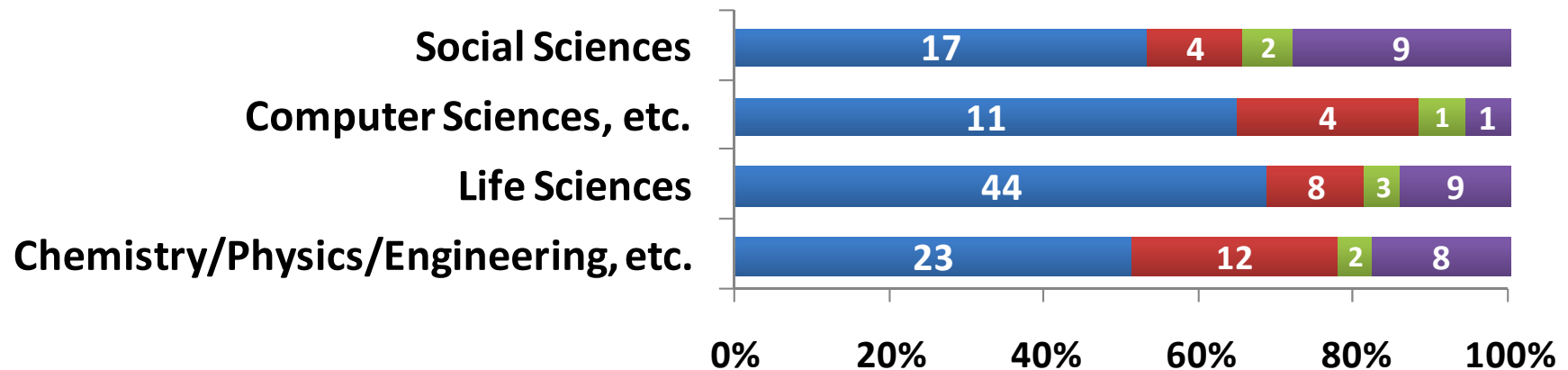


80% chose to reject even review articles if they had "cut and paste" copying, or accept if the copied material had been rewritten in the author's own words.



Q20. Acceptable % similarity index in review papers (SV1, n=161)

■ Under 35% ■ 35%-50% ■ Over 50% ■ Other



1. In Q20: 62% of respondents said that the acceptable OSI in a review article would be under 35%.
2. There were no marked differences in disciplines.



Q21: Percentage of papers rejected because of plagiarism in Western and Non-Western (sv1 & sv2)

Journals	Rejected specifically on account of plagiarism (mean, %)	<i>n</i>
Western	3.5%	102
Non-Western	11.0%	51
All	5.9%	153

Q21. In your own journal(s) and hence subject area, approximately what percentage of papers you receive are rejected specifically on account of plagiarism?



Q22. Willingness to refer to use of CrossCheck in "Instructions to Authors" (SV1)

Response	All (<i>n</i> =148)
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 counter-productive	19%

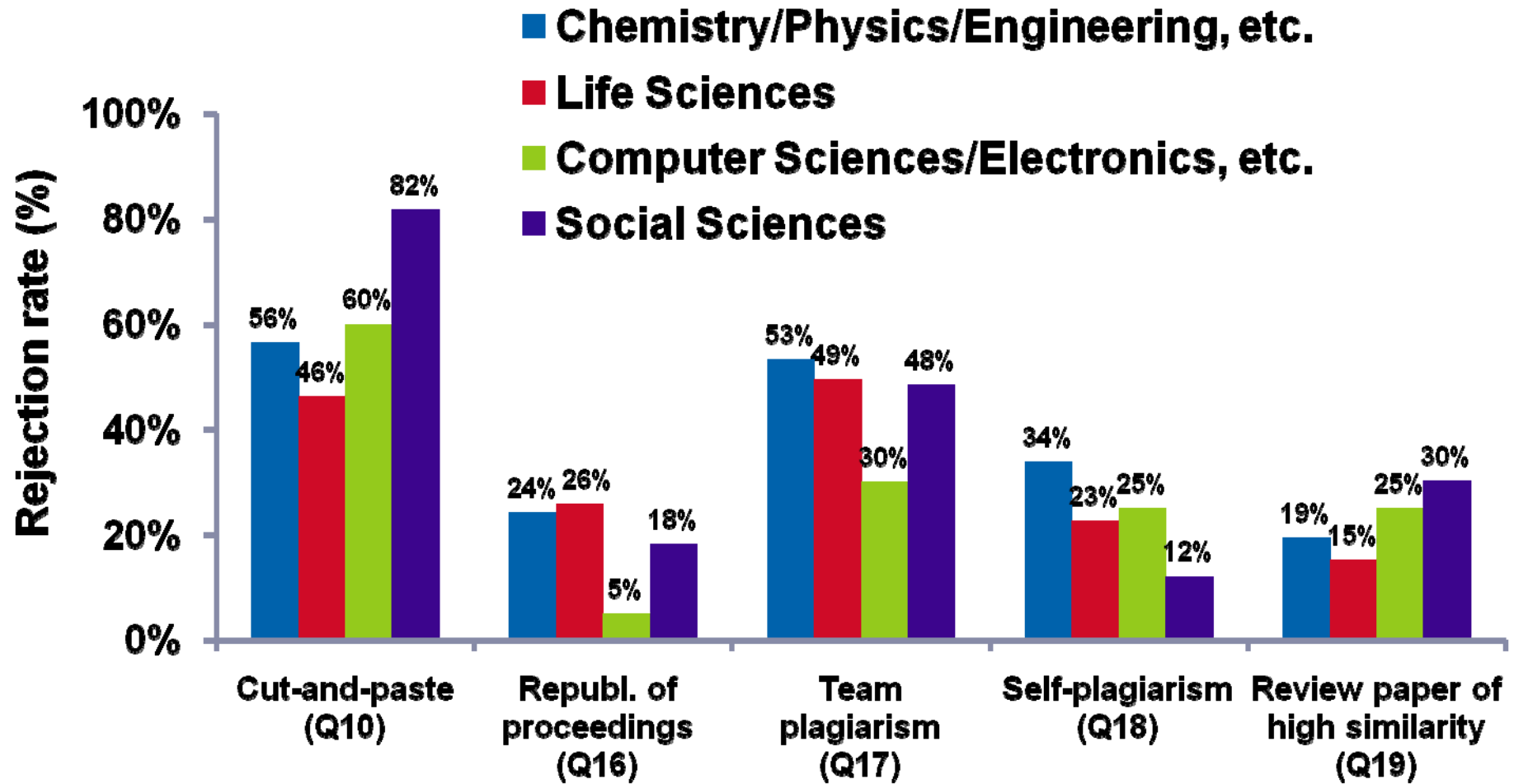


Next charts will illustrate the differences in reactions from both disciplines and languages to 5 plagiarism problems

- 1. Cut and paste**
- 2. Republication of proceedings papers**
- 3. Team plagiarism**
- 4. Self-plagiarism**
- 5. Heavy use of copied material in review papers**



Discussion by discipline: Disciplinary differences in REJECTION rates to 5 key questions



Questions surveyed both in SV1 and SV2



Discussion by discipline: Disciplinary differences in REJECTION rates to 5 key questions

Social Sciences show the **lowest tolerance**, with 82% and 30% rejection rates, respectively to **cut-paste** & much **copied material in review papers**, but to **self-plagiarism** show a wide 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. The writing process is a creative process, so social science editors express the lowest tolerance here. Otherwise, Further discussion of self-plagiarism may be needed in the future."

Computer science shows the **lowest rejection rate** in **replication of proceedings papers**, and **team plagiarism**, with 5% and 30%, respectively. Because this subject mainly depends on the updating of new technologies and team cooperation, there are more conference proceedings publications. Establishing a new policy is most urgent, owing to changing publishing modes and ethics."

The attitude toward the five questions in the Chemistry/Physics/Engineering and Life Sciences seem to be very similar. However, heavy use of copied material in life science review papers shows a little less tolerance"



Discussion by Anglophone & non-Anglophone: Mainstream & Small Differences on 5 key questions

Fig (a) will show us the mainstream views from Anglophone & non-Anglophone respondents,

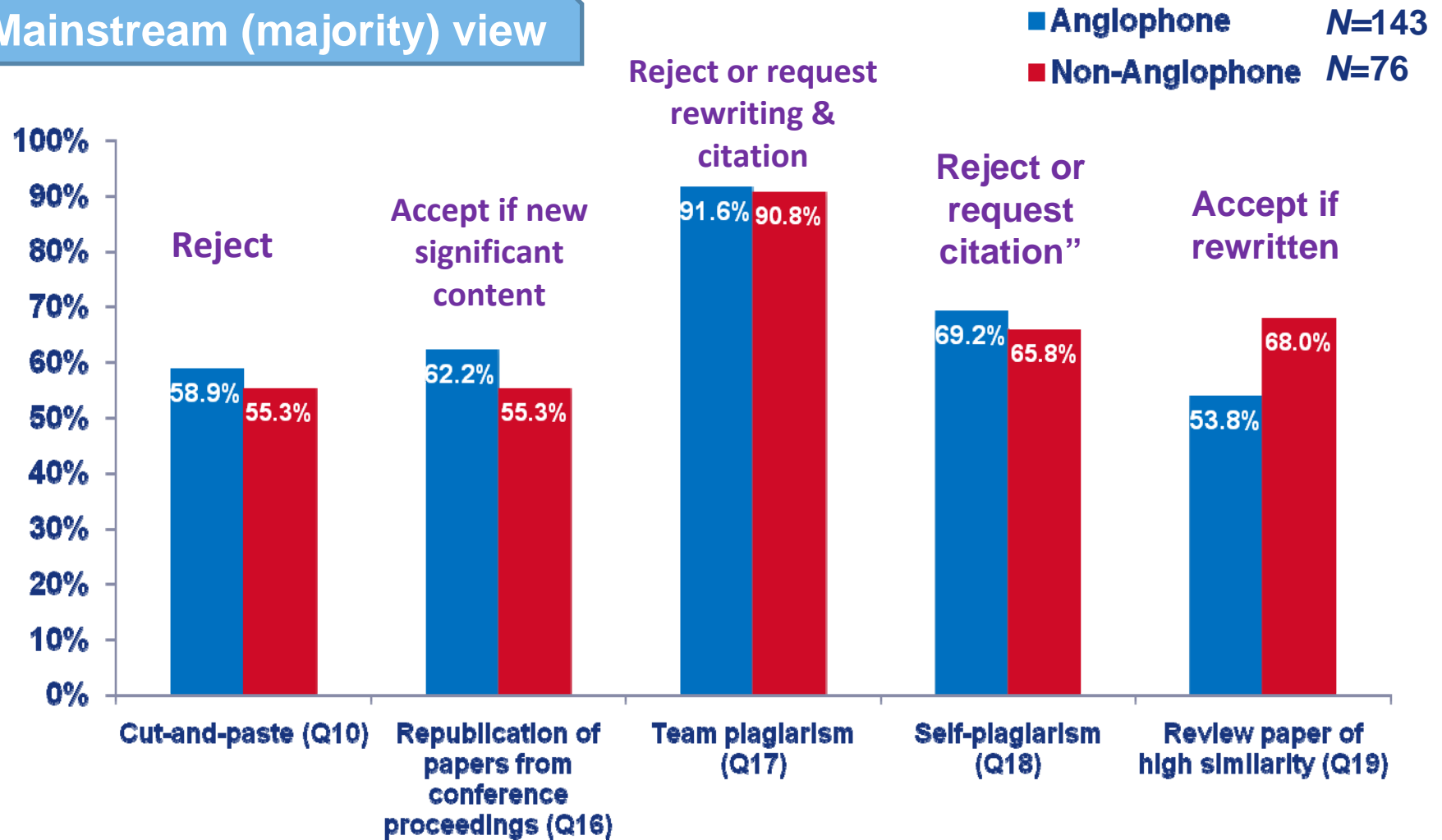
Which proves global editors have expressed a strong mainstream view about ethical standards.



(a) Mainstream (majority) view

between Anglophone & non-Anglophone respondents on 5 key questions

(a) Mainstream (majority) view

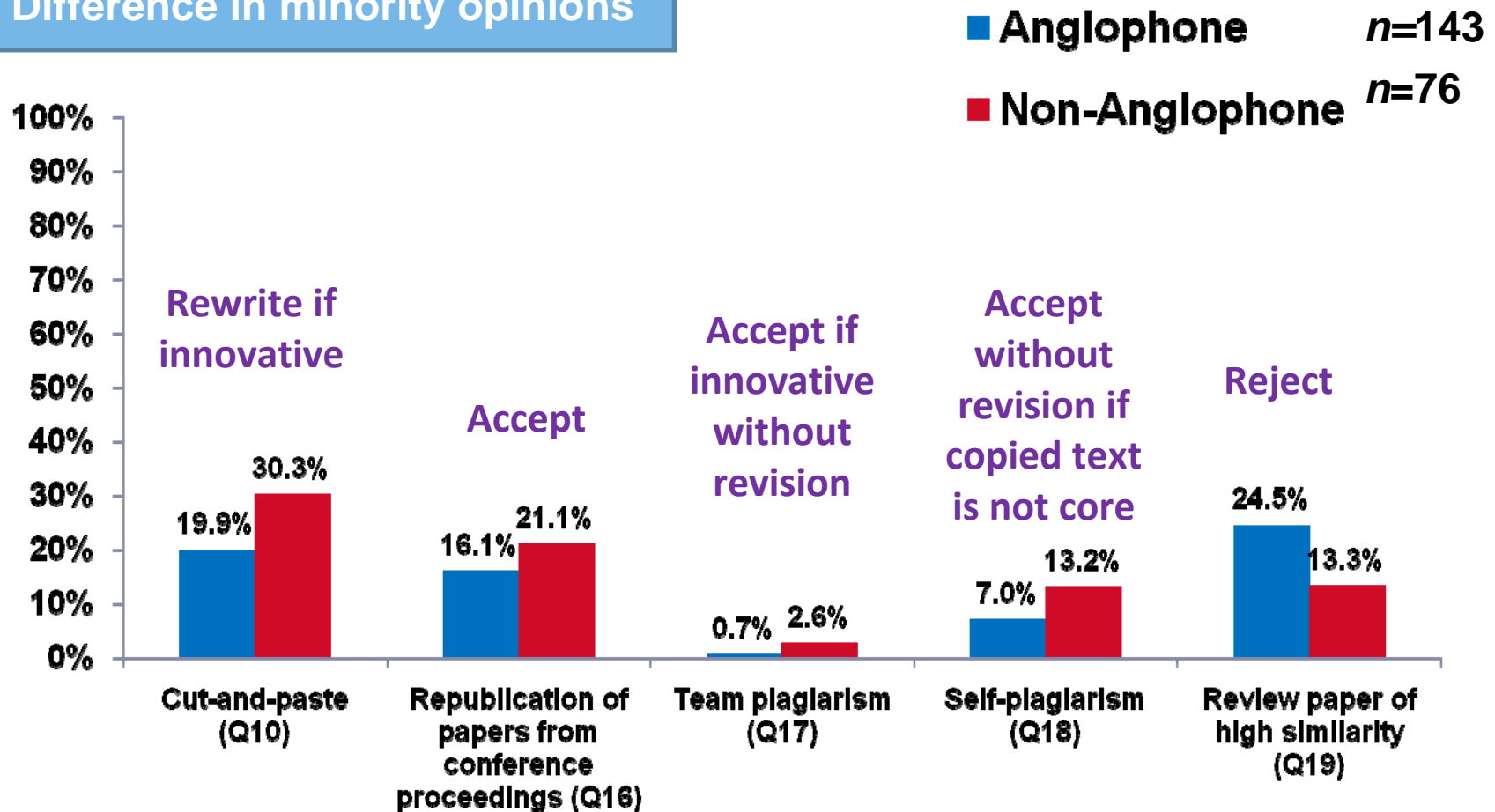




(b) Small Differences

between Anglophone & non-Anglophone respondents on the 5 key questions

(b) Difference in minority opinions



Questions surveyed both in SV1 and SV2



Discussion: (b) Small Differences

between Anglophone & non-Anglophone respondents on 5 key questions

Fig (b) shows there are small variations: **the editor attitudes from non-western (non-Anglophone) countries are a little less rigorous than that of western (Anglophone) countries. The editors from non-Anglophone countries show a little high tolerance to 5 questions**

Discussion: These differences may be due to Cultural and Language differences arising from the wide range of social perspectives and stages of national development. As far as we know, copyright law has been well-documented for more than 300 years in western countries, whereas copyright law has been established much more recently in developing countries (for example, in China international copyright law has been in effect only since 1991). So for some authors from developing and non-Anglophone countries, there needs to be time to catch up with both the “historical differences” and “language difference” to lower the incidence of plagiarism.



Conclusions

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



This research has been interesting for us but has been a learning exercise. In the future we hope to learn from this and we welcome any comments and suggestions to help us improve our future work.

Also we would like to express our sincere thanks to all respondents in our survey. And we thank all of experts for their active suggestions and kind help in the survey design, carrying and paper writing.

The survey results have been published :

**Y.H(Helen) Zhang & X.Y. Jia “A survey on the use of CrossCheck for detecting plagiarism in journal articles ”
*Learned Publishing, 25: 292–307, 2012***

<http://alpsp.publisher.ingentaconnect.com/content/alpsp/lp/2012/00000025/00000004/art00008>

Thank You !



www.zju.edu.cn/jzus

jzus@zju.edu.cn