

Should author self-citations be excluded from citation-based research evaluation? Perspective from in-text citation functions

Dangzhi ZHAO & Alicia CAPPELLO

INTRODUCTION

Author self-citation can, in theory, be considered a natural part of citation behaviour as science builds on previous research by the scientific community, including the citing authors themselves. Self-citations then become an important source of information for studying scholarly communication patterns using citation analysis (Glanzel et al., 2006). In practice, however, author self-citations are often considered as problematic and consequently excluded from evaluative citation analysis (MacRoberts & MacRoberts, 1989; Aksnes, 2003). They tend to be excluded because self-citations are often not considered to represent realistic impacts on the scientific community and because authors can manipulate self-citations in their favour. Manipulating self-citations could potentially happen more often with the increased use of citation analysis in research evaluation exercises and funding or promotion decisions, both affecting the work and lives of researchers. The assumption here is that self-citations do not conform to the norms that govern citation behavior in science (i.e., citing for giving due credit) and therefore play a nonessential role in the citing papers.

Both self-citations and studies on self-citations prevail, but most studies have been aggregate in nature and tend to centre around self-citation rates and how those differs across research fields. Some studies examined the difference that self-citations make in citation indicators such as the h-index (Brown, 2009; Gianoli & Molina-Montenegro, 2009; Bartneck & Kokkelmans, 2011; Ferrara & Romero, 2013), or the impact of self-citations on collaboration (Glanzel & Thijs, 2004; Hellsten et al., 2007; Lin & Huang, 2012). Except for Glanzel, Thijs, & Schlemmer (2004), we have not seen studies on whether the above-mentioned assumption holds, or the degree to which the assumption holds. Glanzel and colleagues studied the regularities of self-citations, and proved through mathematical models that self-citations are an organic part of the citation process and do not need to be removed from citation analysis, at least at the macro-level.

The present study fills this gap by examining individual self-citation occurrences in the citing articles to determine if self-citations function in a nonessential role more or less than foreign citations. Results should contribute to a more precise interpretation and treatment of self-citations in evaluative citation analysis.

METHODOLOGY

Data collection details, and the care we took to reduce the impact of limitations associated with the data collection approach, can be found in Zhao, Cappello and Johnston (2017). We provide an overview here.

We collected all research articles from a single issue of the *Journal of the Association for Information Science and Technology* (JASIST) 2016, volume 67 issue 1. There were 14 articles in total, and they contain 1,473 in-text citations. We coded all these individual citation occurrences as to their function

39 based on the context in which they appear. We noted whether a cited reference was a self-citation
40 defined as a reference that shares at least one author with the citing paper.

41 The coding scheme we used has five categories: Applied, Contrastive, Supportive, Reviewed, and
42 Perfunctory (Tabatabaei, 2013, pp. 153-176). We can reasonably consider citations in the first three
43 categories (i.e., Applied, Contrastive, or Supportive) as having substantial influence on the citing paper
44 and those in the last two categories (i.e., Reviewed or Perfunctory) as nonessential citations.

45 The 14 source articles were coded in random order. Each of the in-text citations were classified into
46 one of the above five functional categories by two coders, independently. If we consider both
47 perfunctory and reviewed citations as a single category labeled “nonessential,” the inter-coder
48 reliability was 85%.

49 Other details of each in-text citation occurrence were also recorded, including the location of the in-
50 text citation within the source article (e.g., Introduction, Methodology, etc.), and how frequently each
51 in-text citation appeared within the source article. Data was manually cleaned and then analyzed
52 according to the function, location, and frequency.

53 As a case study of the Library and Information Science (LIS) field, as represented by a single issue of
54 JASIST, the present study suffers from the limitation of scalability and generalizability. Future studies
55 are required to verify if the patterns found here apply to other research fields.

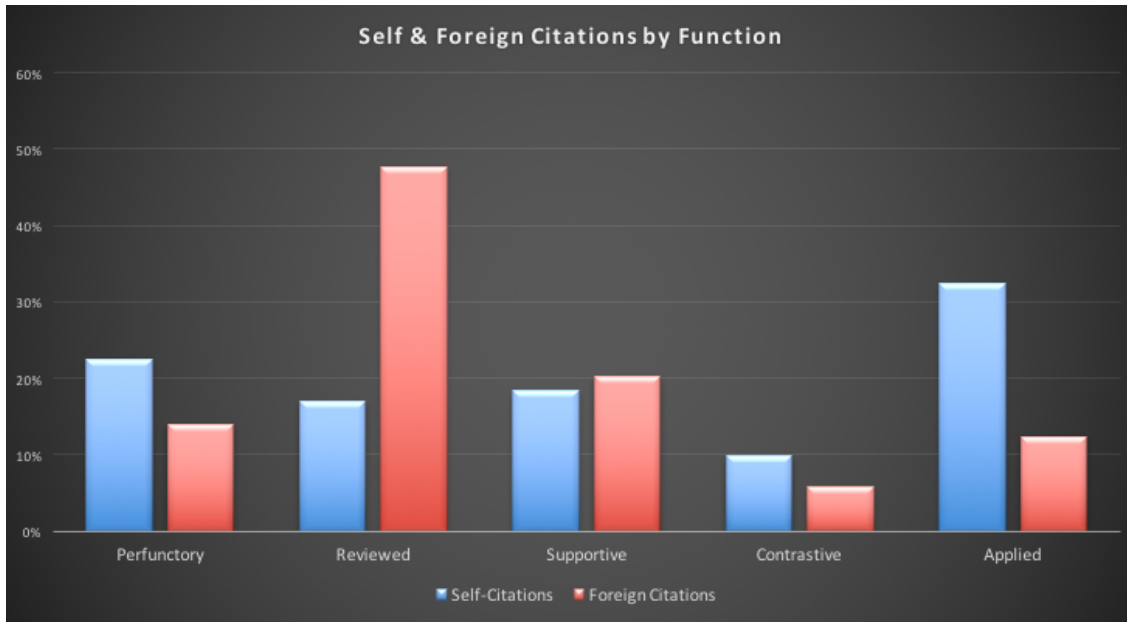
56

57 **RESULTS**

58

59 ***Function***

60 Figure 1 compares self-citations and foreign citations in terms of the percentage of *unique* citations in
61 the five functional categories. Each item listed in the reference list at the end of each article is
62 considered a unique citation, regardless of how many times was cited within the citing article. All
63 instances of that unique citation were coded by function, and the function that had the highest impact
64 was then assigned to that unique citation. For example, the citation “Shenton and Dixon 2003” appears
65 four times as Contrastive, one time as Supportive, and one time as Reviewed in citing article #3. The
66 Contrastive function has the highest impact among the three functional categories, and was thus
67 assigned to this citation in article #3.



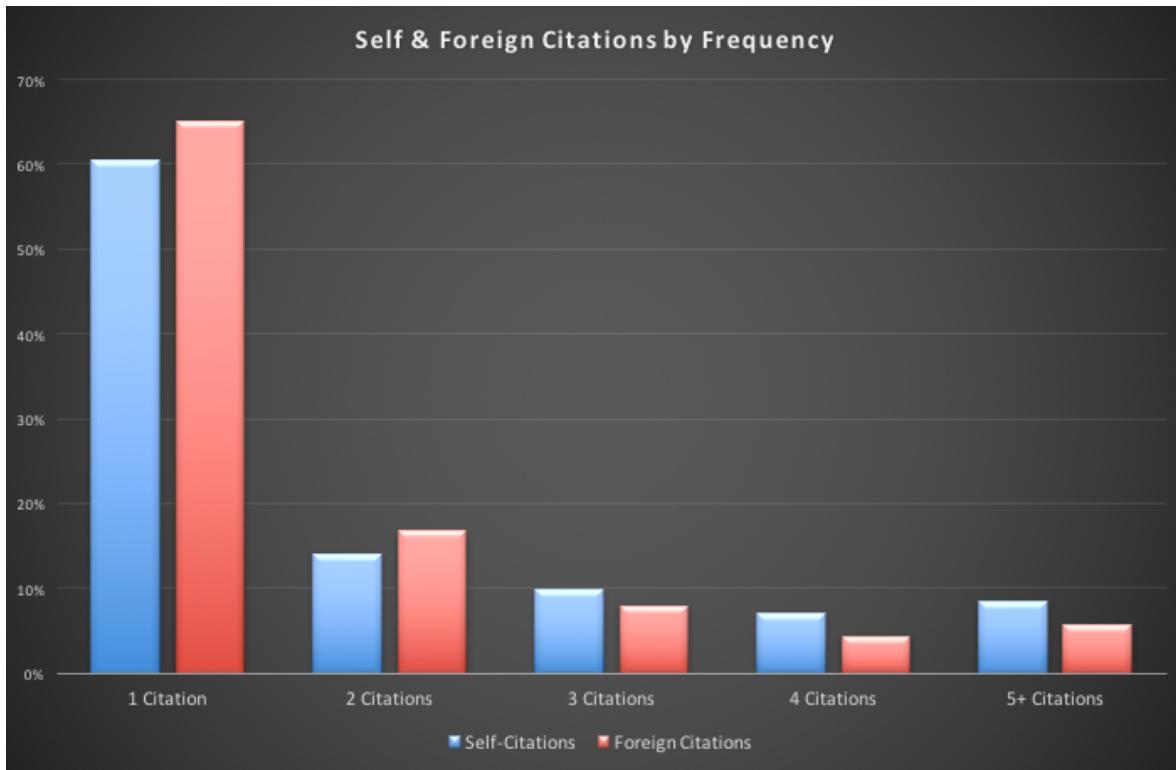
68

69 Figure 1: Self and Foreign Unique Citations by Function

70 It is interesting to note that the percentage of self-citations that represent a substantial impact on
 71 citing papers (i.e., sum of Supportive, Contrastive, and Applied citations) is much higher than the
 72 percentage of foreign citations for the same functional categories (more than 60% vs. less than 40%).
 73 About one-third (32%) of self-citations are in the highest impact function (Applied), which is more than
 74 double that of foreign citations in the same function (~12%). It appears that, in general, self-citations
 75 are more likely to be Applied citations than any other citation function, and are less likely to function
 76 as nonessential citations compared to foreign citations. This result suggests that self-citations should
 77 not be discounted in citation analysis as many previous studies have promoted, and should in fact be
 78 given more weight than foreign citations in weighted citation analysis.

79 **Frequency**

80 Each unique citation was put into one of five categories of citation frequency: uni-citation, 2 citations,
 81 3 citations, 4 citations, and 5+ citations. For example, the above-mentioned “Shenton and Dixon 2003”
 82 citation appears six times in total in citing article #3, and was thus assigned to the frequency category
 83 “5+ citations.” Figure 2 compares self-citations and foreign citations in terms of the percentage of
 84 unique citations in these five frequency categories.



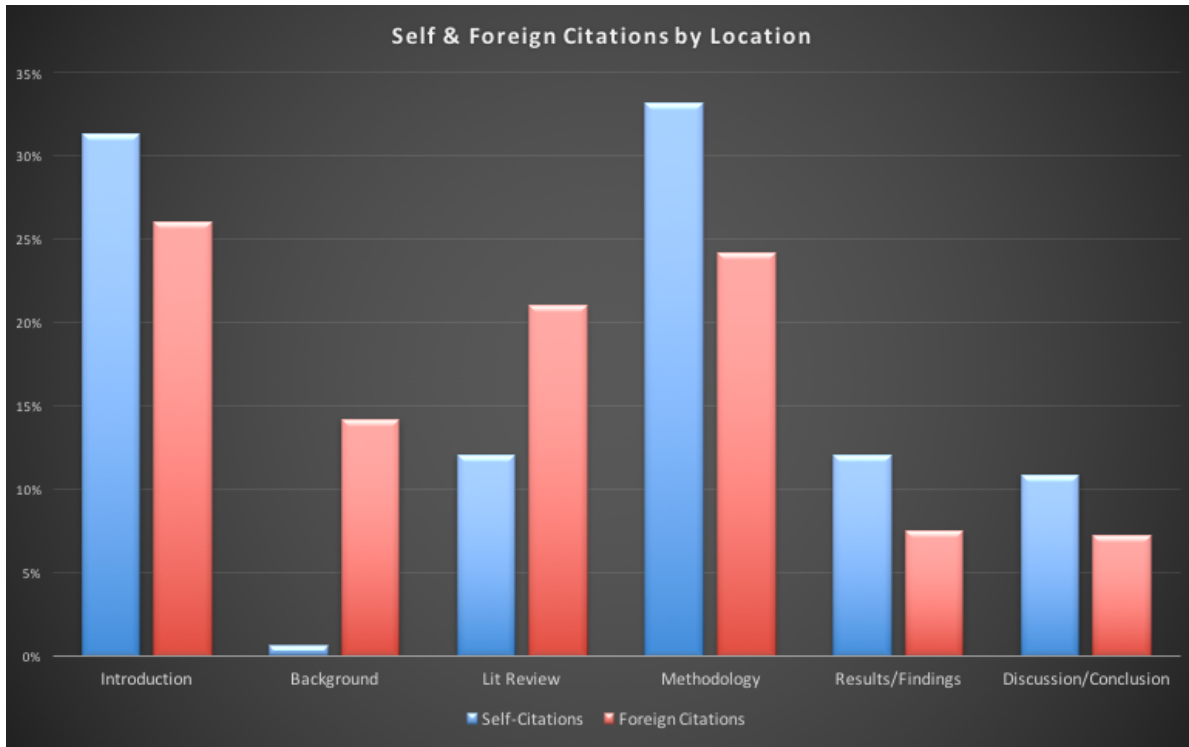
85

86 Figure 2: Self and Foreign Unique Citations by Frequency

87 Figure 2 shows that both self-citations and foreign citations appear mainly as uni-citations within the
 88 citing paper. It is interesting to note that the percentages of self-citations in the higher frequency
 89 categories (i.e., 3, 4, and 5+) are higher than those of foreign citations. This result indicates that self-
 90 citations are less likely to be nonessential than foreign citations because it has been found that the
 91 likelihood of a citation being nonessential (i.e., perfunctory or reviewed) decreases with the frequency
 92 at which it appears in the citing text (Zhao, Cappello & Johnston, 2017).

93 **Location**

94 Next we examine the percentage of self-citations and foreign citations in specific locations (e.g.,
95 Introduction, Methodology, etc.) within the citing article (Figure 3). All in-text citation occurrences in a
96 citing paper were included in the calculation, as each occurrence of the citation may appear in a
97 different section. For example, the above-mentioned “Shenton and Dixon 2003” citation appears a
98 total of six times in citing article #3, and was therefore counted six times.



99

100 Figure 3: Self and Foreign In-Text Citations by Location

101 As seen from Figure 3, a much lower percentage of self-citations than foreign citations appear in the
102 Background and Literature Review sections, and a much higher percentage of self-citations than
103 foreign citations appear in the Methodology, Results/Findings, and Discussion/Conclusion sections.
104 This result supports findings from analyses above that self-citations are less likely to be nonessential
105 than foreign citations. This claim is informed by findings from a previous study that nonessential
106 citations dominate the Background and Literature Review sections at 97% and 93% respectively, but
107 account for only 40% or less of all in-text citation occurrences in each of the other sections (Zhao,
108 Cappello, & Johnston, 2017).

109 **DISCUSSION AND CONCLUSION**

110 Self-citations have long been noted as a problem in citation analysis and are often excluded from the
111 analyses based on the notion that self-citations may be included for egoistic or self-serving reasons.
112 The present study, however, found that self-citations are less likely to function as nonessential
113 citations than foreign citations, leading to the conclusion that self-citations should not be discounted in
114 citation analysis, and should in fact be given more weight than foreign citations in weighted citation
115 analysis.

116 Egoistic and self-serving citations do exist (Edge, 1979; Cronin, 1984), and may increase with the
117 increasing use of citation-based metrics in research evaluation (Glanzel et al., 2006). Solving this

118 problem may require some fundamental changes in the scholarly communication system, changes that
119 have been made possible and desirable by digital environments. In the meantime, some feasible
120 changes to editorial policy may help. For example, journals could perhaps require that articles list
121 essential and nonessential citations differently and/or separately, and ask authors to explain the
122 connection of self-citations to the citing articles, as some already do for author contributions to multi-
123 authored articles.

124 **ACKNOWLEDGMENT**

125 We would like to thank Lucinda Johnston for her help with data collection.

126 **REFERENCES**

- 127 Aksnes, D.W. (2003). A macro study of self-citation. *Scientometrics*, 56(2), 235-246.
- 128 Bartneck, C., & Kokkermans, S. (2011). Detecting h-index manipulation through self-citation analysis.
129 *Scientometrics*, 87, 85-98.
- 130 Brown, R.J.C. (2009). A simple method for excluding self-citation from the h-index: the b-index. *Online*
131 *Information Review*, 33(6), 1129-1136.
- 132 Cronin, B. (1984). *The citation process. The role and significance of citations in scientific*
133 *communication*. London: Taylor Graham.
- 134 Edge, D. (1979). Quantitative measures of communication in science: A critical review. *History of*
135 *Science Cambridge*, 17(36), 102-134.
- 136 Ferrara, E., & Romero, A.E. (2013). Scientific Impact Evaluation and the Effect of Self-Citations:
137 Mitigating the Bias by Discounting the h-index. *Journal of the American Society for Information*
138 *Science and Technology*, 64(11), 2332-2339.
- 139 Gianoli, E., & Molina-Montenegro, M.A. (2009). Insights into the Relationship Between the h-index and
140 Self-Citations. *Journal of the American Society for Information Science and Technology*, 60(6),
141 1283-1285.
- 142 Glanzel, W., Debackere, K., Thijs, B., & Schubert, A. (2006). A concise review on the role of author self-
143 citations in information science, bibliometrics and science policy. *Scientometrics*, 67(2), 263-
144 277.
- 145 Glanzel, W., & Thijs, B. (2004). Does co-authorship inflate the share of self-citations? *Scientometrics*,
146 61(3), 395-404.
- 147 Glanzel, W., Thijs, B., & Schlemmer, B. (2004). A bibliometric approach to the role of author self-
148 citations in scientific communication. *Scientometrics*, 59(1), 63-77.
- 149 Hellsten, I., Lambiotte, R., Scharnhorst, A., & Audloos, M. (2007). Self-citations, co-authorships and
150 keywords: A new approach to scientists' field mobility? *Scientometrics*, 72(3), 469-486.
- 151 Lin, W-Y.C., & Huang, M-H. (2012). The relationship between co-authorship, currency of references and
152 author self-citations. *Scientometrics*, 90, 343-360.
- 153 MacRoberts, M.H., & MacRoberts, B.R. (1989). Problems of citation analysis: A critical review. *Journal*
154 *of the American Society for Information Science*, 40(5), 342-349.
- 155 Tabatabaei, N. (2013). *Contribution of information science to other disciplines as reflected in citation*
156 *contexts of highly cited JASIST papers*. Montreal. (McGill University P.hD. dissertation).
- 157 Zhao, D., Cappello, A., & Johnston, L. (2017). Functions of Uni- and Multi-citations: Implications for
158 Weighted Citation Analysis. *Journal of Data and Information Science*, 2(1), 51-69.