

*promoting access to White Rose research papers*



**Universities of Leeds, Sheffield and York**  
**<http://eprints.whiterose.ac.uk/>**

---

This is an author produced version of a paper published in **Journal of Documentation**.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/76253>

---

**Published paper**

Willett, P. (2013) *Readers' perceptions of authors' citation behaviour*. Journal of Documentation, 69 (1). 145 - 156. ISSN 0022-0418  
<http://dx.doi.org/10.1108/00220411311295360>

---

# Readers' Perceptions of Authors' Citation Behaviour

## 1. Introduction

The introduction of the ISI citation databases that became the Web of Science (and subsequently of other citation databases) has driven the development of many types of bibliometric research, as reviewed by Baird and Oppenheim (1994), Bar-Ilan (2008), Garfield (1979) and Nicolaisen (2007) *inter alia*. Here, we focus on studies of citing behaviour, specifically the reasons why authors make citations to the previous literature when they are preparing articles, books and other types of scholarly publication (Bornmann and Daniel, 2006; Cronin, 1984; Liu, 1993). Such studies are often referred to as *citation context analysis*, a technique that seeks “..to illuminate the inter-document relationship implied by the presence of citations by devising a classification or taxonomy based on an analysis of the text surrounding the citations..” (Bornmann and Daniel, 2006). Ritchie (2008) discusses the relationship between citation context analysis, and the related, partially overlapping approaches of citation content analysis, citation classification and content citation analysis.

Two rather different approaches have been used to investigate the reasons for citing a particular reference. In the first, direct approach, authors are asked why they included each of the citations in their publications. Such studies can be problematic since it may be difficult to obtain an author's co-operation or an author may not be able to remember the rationale for a citation if the publication is at all old (Wilkinson *et al.*, 2003). In the second, indirect approach, the rationale is obtained from careful reading by a non-author of the text that contains and relates to the citation under study. The latter, reader-based approach can be far more convenient and less intrusive than approaching the original author to provide the rationale for citing, but there are at least two consequent problems: the reader must have sufficient subject knowledge to be able to understand the contexts of the citations that are being studied; and, importantly, there is an implicit assumption that the contexts are written in such a way that the reader is able to recognise the original purpose of the citation when the article was written (Borgman and Furner, 2002).

Bornmann and Daniel (2006) note that there has been no attempt to link citer motivation and citation context analyses by integrating the perspectives provided by authors and by readers. This paper attempts such an integration by comparing the reasons for citation suggested by the authors of a set of ten journal articles in librarianship and information science with the reasons for citation suggested by readers of that same set of articles. If it can be shown that there is a close measure of agreement between the results of the two approaches then this would validate the use of reader-based citation context analysis as an effective proxy for the citation behaviour of the original authors of articles. If, however, there is only a limited agreement then this would raise questions as to the appropriateness of this approach for the study of citation behaviour.

## **2. Methods**

### ***2.1 Basic procedure***

The study has been designed to overcome, or at least to alleviate, the problems that have been noted in the Introduction. First, the articles that have been used had all been published only a few months prior to this study being carried out, and the reasons for citation should hence have been still relatively fresh in the minds of the authors. Specifically, the study used ten articles from an issue of volume 63 of the journal *Aslib Proceedings* that highlighted the work of the Information School at the University of Sheffield and that were all written by members of the academic staff of the School (and hence colleagues of the author of this article). The authors, title and page numbers for each of the ten articles are listed in Appendix 1. Second, the readers were members of the academic staff, research staff and research students in the School, who hence all had a background in librarianship and information science sufficient for detailed understanding of the article(s) on which they were making judgements of the author's motivations for citing references.

The author (or one of the authors when more than one member of the School's staff was involved) of each of the *Aslib Proceedings* articles was asked to select ten of the citations that they considered of greatest importance in the text of their article, and then to assign a reason to each of these citations using the modified Harwood classification scheme. The scheme, which is discussed in detail below, has a total of 18 possible reasons for an author to make a citation, and hence each author was asked to provide a marked-up copy of their article with

the ten chosen citations highlighted at the point that they occurred in the text, and a separate list of ten integers, each in the range 1-18, with the  $i$ -th integer denoting the *author-reason* for the  $i$ -th chosen citation. A reader was then given a copy of the article with the chosen citations highlighted, and asked to classify these citations using the same classification scheme, thus yielding a further ten-integer set of reasons. Each reader was asked to choose for analysis an article for which they felt they had a sufficient degree of subject knowledge to make meaningful judgements of the reasons for citation, with several of them providing such sets of *reader-reasons* for more than just a single such article.

The similarity between an author's set of reasons and a reader's set of reasons was then computed using the Dice coefficient, a simple association coefficient. Given two lists containing  $A$  and  $B$  integers,  $C$  of which are common to the two lists, the Dice coefficient is  $2 \times C / (A + B)$ . Here,  $A = B = 10$  and hence the coefficient is simply  $C / 10$ , with values of zero and one indicating no reasons in common and all reasons in common, respectively. The use of the Dice coefficient in the present context was suggested by its analogous use by Rolling (1981) to measure the extent to which two indexers select the same index terms to characterise a document's content. A sub-set of the total citations in each article was used since this would maximise the chance that authors and readers would be willing to contribute to the study. Additionally, the Dice coefficient (and analogous association coefficients such as the Tanimoto coefficient or the cosine coefficient) normalises the number of common features by the lengths of the lists: the use of a fixed list-length removes this source of variation in the computed coefficient values.

A total of 24 individuals provided sets of reader-reasons on one or more of the ten articles, with each article being read by between three and six different readers. In all, the individuals provided a total of 45 sets of reader-reasons, and these 45 sets provided the basis for the analyses in the Results and Discussion section below.

## **2.2 *Classifications of citer motivations***

Studies of citation behaviour, whether for citation context analysis or for the other types of study discussed by Ritchie (2008), often make use of a classification of reasons for citation,

e.g., paying homage to pioneers, giving credit for related work, or providing background reading. These three reasons are just some of those suggested in what is probably the first such classification (Garfield, 1965). When an individual publication is being analysed, a classification scheme may be designed that is specific to that publication: examples of this approach include studies of books on software engineering (McCain and Salvucci, 2006) and on strategic management (Anderson, 2006) and of articles on molecular biology (Ahmed *et al.*, 2004; McCain and Turner, 1989), organizational theory (Anderson and Sun, 2010; Lounsbury and Carberry, 2005; Mizruchi and Fein, 1999) and neuropharmacology and the sociology of science (Cozzens, 1985). Other classification schemes have been developed for analyzing citations to the works of an individual author (Brittain, 2000) and to an individual journal (Spiegel-Rosing, 1977) or, most commonly, for use across the whole range of subjects and types of publication (Bonzi and Snyder, 1991; Cano, 1989; Case and Higgins, 2000; Chubin and Moitra, 1975; Meho and Sonnenwald, 2000; Moravesik and Murugesan, 1975; Oppenheim and Renn, 1978; Peritz, 1983; Shadish *et al.*, 1995; Vinkler, 1987).

Different schemes contain very different numbers of reasons for citation. For example, the scheme used by Case and Miller (2010) in a recent study of the citing behaviour of bibliometricians contained 33 possible reasons, whereas that used by Meho and Sonnenwald (2000) to study publications in the field of Kurdish studies contained just six reasons. The nature of the reasons also varies, although to a lesser extent than might seem apparent on first glance since careful inspection of the studies cited above reveals a considerable degree of overlap. Overviews of citation behaviour have been presented by Bornmann and Daniel (2006), Liu (1993) and White (2004), and there have also been reports of the development of analogous classification schemes to facilitate studies of the reasons for the linking of pairs of websites (Chu, 2005; Wilkinson *et al.*, 2003).

### **2.3 *The Harwood classification scheme***

From the many classifications of citations that are available, we have chosen to base our work on the scheme described by Harwood (2009). This contains eleven main types of citation, with several of these being sub-divided to give a total of 24 different reasons.

Harwood's scheme was developed as a result of extended interviews with the authors of six computer science articles and six sociology articles, and thus reflects directly the authors' reasons for citing. In the present study, the scheme is to be used both by authors and by

readers, and some minor modifications to it have been required. Most obviously, the fact that the scheme was devised solely on the basis of authors' comments means that some of the reasons are unlikely to be recognised by readers. For example, one section of the classification, referred to by Harwood as Advertising, described the use of citations specifically to alert readers to earlier work by the author or by others. This is an entirely reasonable motive for citing but it might be difficult for a reader to differentiate this from the section of the classification, called Building, which relates to the use of citation to identify earlier publications, either by the author or by others, that act as building blocks for the current publication. Again, the use of a citation that is intended by the author to display ability to conduct future research (one of the reasons listed in the Competence section of the classification) might well be indistinguishable to a reader from a citation that is intended by the author to map out the author's future research plans (the reason listed in the Future section). The removal of these potentially ambiguous reasons means that we are likely to increase the degree of commonality between sets of author and reader judgements, as is the addition of a 'cover-all' reason – Other - for any citation that cannot be assigned to any of the other reasons. In all, the resulting classification contains the 18 reasons listed in Appendix II. It will be seen that the reasons have been organised in ten broader groups, thus providing an alternative, less specific set of reasons for determining the degree of author-reader overlap (*vide infra*).

### **3. Results and Discussion**

Ten sets of author-reasons and 45 sets of reader-reasons were collected. The percentage frequency distributions for the resulting 100 author-reasons and 450 reader-reasons are shown in Figure 1. There appears to be no significant difference between the two distributions and this is confirmed by a Kolmogorov-Smirnov test ( $p = 0.766$  in a two-tailed test using the Aston University Statistical Software at [http://www.wessa.net/rwasp\\_Reddy-Moores%20K-S%20Test.wasp/](http://www.wessa.net/rwasp_Reddy-Moores%20K-S%20Test.wasp/)). However, this broadly comparable usage of the various reasons does not translate into high author-reader similarities. These similarities are shown in Table 1 where each element in a row of the table contains the number of author-reader comparisons with a particular similarity value. We consider first the results in column A, which represents the use of the 18-reason Harwood classification as listed in Appendix II. Thus, for example, the first two elements in this column show that there was not a single

reason in common for 7 of the 45 author-reader comparisons, and that there was just a single reason (i.e. a Dice similarity of 0.1) in common for 12 of these comparisons.

Table 1 and Figures 1-2 near here

The distribution of the 45 values of the Dice coefficient in column A is shown in Figure 2. A basic assumption of citation context analysis is that readers are able to perceive the authors' reasons for citing, which would result in high mean values, ideally values of unity, for the Dice coefficient in our experiment. However, it will be seen that the distribution is shifted very markedly to the left, with a marked preponderance of low-values (mean = 0.21, standard deviation = 0.17). We have noted previously that between three and six sets of reader-reasons were available for each individual article. Very similar results (mean = 0.22 and standard deviation = 0.16) to those in Figure 2 were obtained using just the first three sets of reader-reasons in those cases where more than three readers had chosen to analyse a particular article, thus removing the possibility that the shifted distribution results from an over-sampling of particular articles.

The preponderance of low similarity values that is observed corresponds to only a small degree of overlap between the authors' reasons for citing particular sources and their readers' subsequent perceptions of those reasons, a finding that immediately raises significant questions as to the appropriateness of citation context analysis as a means of investigating citation behaviour. That said, the results obtained here seem to be at least better than would be obtained if authors and readers assigned reasons for citation at random. With 18 reasons available for assignment, any one of them has an *a priori* probability of 1/18 (i.e., approximately 0.0556) of being chosen as the reason for a particular citation. A binomial test then shows that one would expect to get no matching reasons at all in over one-half of all the author-reader comparisons; in fact, Table 1 shows that such zero matches comprised less than 16% of the author-reader comparisons, and Figure 1 shows that individual reasons are used to very different extents. Thus, while the observed author-reader similarities are markedly different from what be expected by chance, they are also markedly less than would be expected if readers were able consistently to perceive authors' reasons for citing. If the latter was indeed the case then one would expect many reader-author similarities of 1.0 but column A in Table 1 shows that this never occurred in our sample; indeed, even values of 0.5 were quite rare.

Inspection of the reasons in Appendix II shows that some of them are quite specific, and a greater degree of overlap might hence be expected if a less specific matching criterion was adopted. This has been done in two ways. First, as described in the Methods section, the 18 reasons have been grouped into ten classes (Signposting, Supporting, Credit etc.) as shown in Appendix II. The original sets of author-reasons and reader-reasons, based on 18 reasons, were converted so that they were based on the ten broader reasons, hence increasing the potential overlap. For example, an author assignment of reason-8 to a citation would enable a match with a reader assignment of reason-10 since they are both examples of the class named Position. Second, when the authors and readers were asked to suggest the most appropriate reason for a particular citation, they were also allowed, if they so wished, to suggest one or more alternative reasons from amongst the 18 available. For example, an author assignment of reason-3 to a citation as an alternative to reason-9 would enable a match with a reader assignment of the former reason. The resulting distributions of coefficient values are shown in the columns headed B (10 reasons) and C (18 reasons but with alternatives) in Table 1. There is clearly some increase in the similarities with the mean (standard deviation) values being 0.28 (0.18) and 0.43 (0.20) for columns B and C respectively (with one author-reader comparison achieving the limiting value of 1.0 when alternatives are included); however the values still indicate a substantial level of mis-match between author and reader perceptions.

The focus of this paper is a comparison of author-reasons and reader-reasons and the results presented above show that there is only a limited level of agreement between these two sets of reasons. A fair measure of disagreement was also noted when comparing the reasons suggested by different readers. For example, one of the articles was assessed by six different readers, with the six author-reader similarities being 0.2 (four occurrences), 0.3 (one) and 0.4 (one). The 15 distinct reader-reader similarities for this article were 0.0 (three occurrences), 0.2 (three), 0.2 (two), 0.3 (five), 0.4 (one) and 0.5 (one), from which it may be concluded that different readers understand the contexts of the same citations in very different ways.

It is important to emphasise the limitations of this study. First, the use of a single classification scheme from amongst the many that are available (*vide supra*). However, we have used the Harwood scheme in three different ways, with little effect on the overall conclusion of a low level of author-reader similarity, and hence have no reason to believe that substantially different results would have been obtained if an alternative scheme had been



employed. Second, the small scale of the exercise with only a few authors and readers working in a single subject area. This has, however, meant that we have been able to include authors only a short time after the writing of their articles and readers who have a high degree of subject knowledge, thus alleviating two of the problems in citation-behaviour studies noted in the introduction to this paper. Third, the use of the author's colleagues means that there may have been a greater degree of shared understanding than would have been the case if the authors and readers had come from different organizations and different academic backgrounds. That said, such a shared understanding might have tended to increase the similarity values over those that would have been obtained if different organizations had been involved, i.e., the values would have been even lower than was the case here.

We hence conclude that reader-reasons and author-reasons for citing particular references are typically so different that it would be unwise to use the former as a proxy for the latter, thus confirming the suggestion of Case and Miller (2010) that readers "...may make judgments that differ a great deal from the intentions of the original author...". The lack of agreement here is, perhaps, hardly surprising when one considers that marked levels of inter-individual disagreement have been found in several analogous contexts, such as the assignment of indexing terms (Markey, 1984), the creation of links in hypertext systems (Ellis *et al.*, 1994) and the selection of search strategies (Iivonen, 1995).

#### **4. Conclusions**

Some previous studies of citation behaviour have made the assumption that reader judgments of the reasons for citation provide an effective proxy for the judgments of the original authors. This small-scale study suggests that there is only limited level of agreement between reader and author judgments; accordingly, while reader-based studies of citation behaviour are relatively easy to carry out they are unlikely to provide accurate insights into the reasons why authors cite material in scholarly texts.

#### **References**

- Ahmed, T., Johnson, B., Oppenheim, C. and Peck, C. (2004) "Highly cited old papers and the reasons they continue to be cited. Part II. The 1953 Watson and Crick article on the structure of DNA", *Scientometrics*, Vol. 61 No. 2, pp. 147-156.
- Anderson, M. H. (2006) "How can we know what we think until we see what we said? A citation and citation context analysis of Karl Weick's The Social Psychology of Organizing", *Organization Studies*, Vol. 27 No. 11, pp. 1675-1692.
- Anderson, M. H. and Sun, P. Y. T. (2010) "What have scholars retrieved from Walsh and Ungson (1991)? A citation context study", *Management Learning*, Vol. 41 No. 2, pp. 131-145.
- Baird, L. M. and Oppenheim, C. (1994) "Do citations matter?", *Journal of Information Science*, Vol. 20 No. 1, pp. 1-15.
- Bar-Ilan, J. (2008) "Informetrics at the start of the 21st century – a review", *Journal of Informetrics*, Vol. 2 No. 1, pp. 1-52.
- Bonzi, S. and Snyder, H. W. (1991) "Motivations for citation: a comparison of self citation and citation to others", *Scientometrics*, Vol. 21 No. 2, pp. 245-254.
- Borgman, C. L. and Furner, J. (2002) "Scholarly communication and bibliometrics", *Annual Review of Information Science and Technology*, Vol. 36, pp. 3-72.
- Bornmann, L. and Daniel, H. (2006) "What do citation counts measure? A review of studies on citing behavior", *Journal of Documentation*, Vol. 64 No. 1, pp. 45-80.
- Brittain, J. M. (2000) "A highly visible scientist - Jack Meadows", *Journal of Information Science*, Vol. 26 No. 4, pp. 267-272.
- Cano, V. (1989) "Citation behavior: classification, utility, and location", *Journal of the American Society for Information Science*, Vol. 40 No. 4, pp. 284-290.
- Case, D. O. and Higgins, G. M. (2000) "How can we investigate citation behavior? A study of reasons for citing literature in communication", *Journal of the American Society for Information Science*, Vol. 51 No. 7, pp. 635-645.
- Case, D. O. and Miller, J. B. (2010) "Do bibliometricians cite differently from other scholars?", *Journal of the American Society for Information Science and Technology*, Vol. 62 No. 3, pp. 421-432.
- Chu, H. (2005) "Taxonomy of inlinked Web entities: what does it imply for webometric research?", *Library and Information Science Research*, Vol. 27 No. 1, pp. 8-27.
- Chubin, D. E. and Moitra, S. D. (1975) "Content analysis of references: adjunct or alternative to citation counting?", *Social Studies of Science*, Vol. 5 No. 4, pp. 423-441.
- Cozzens, S. E. (1985) "Comparing the sciences: citation context analysis of papers from neuropharmacology and the sociology of science", *Social Studies of Science*, Vol. 15 No. 1, pp. 127-153.
- Cronin, B. (1984) *The Citation Process. The Role and Significance of Citations in Scientific Communication*, Taylor Graham, London.
- Ellis, D., Furner-Hines, J. and Willett, P. (1994) "On the creation of hypertext links in full-text documents: measurement of inter-linker consistency. *Journal of Documentation*, Vol. 50 No. 2, pp. 67-98.
- Garfield, E. (1965) "Can citation indexing be automated?", in Stevens, M. E., Giuliano, V. E. and Heilprin, L. B. (eds.), *Statistical Association Methods for Mechanized Documentation*, National Bureau of Standards, Washington, pp. 189-192.

- Garfield, E. (1979) *Citation Indexing -- Its Theory and Application in Science, Technology, and Humanities*, ISI Press, Philadelphia, PA.
- Harwood, N. (2009) "An interview study of the functions of citations in academic writing across two disciplines", *Journal of Pragmatics*, Vol. 41 No. 3, pp. 497-518.
- Iivonen, M. (1995) "Consistency in the selection of search concepts and search terms", *Information Processing and Management*, Vol. 31 No. 2, pp. 173-190.
- Liu, M. (1993) "The complexities of citation practice: a review of citation studies", *Journal of Documentation*, 49, No. 4, pp. 370-408.
- Lounsbury, M. and Carberry, E. J. (2005) "From King to Court Jester? Weber's fall from grace in organizational theory", *Organization Studies*, Vol. 26 No. 4, pp. 501-525.
- Markey, K. (1984) "Inter-indexer consistency tests", *Library and Information Science Research*, Vol. 6 No. 2, pp. 155-177.
- McCain, K. W. and Salvucci, L. J. (2006) "How influential is Brooks' Law? A longitudinal citation context analysis of Frederick Brooks' *The Mythical Man-Month*", *Journal of Information Science*, Vol. 32 No. 3, pp. 277-295.
- McCain, K. W. and Turner, K. (1989) "Citation context analysis and aging patterns of journal articles in molecular genetics", *Scientometrics*, Vol. 17 No. 1/2, pp. 127-163.
- Meho, L. I. and Sonnenwald, D. H. (2000) "Citation ranking versus peer evaluation of senior faculty research performance: a case study of Kurdish scholarship", *Journal of the American Society for Information Science*, Vol. 51 No. 2, pp. 123-138.
- Mizruchi, M. S. and Fein, L. C. (1999) "The social construction of organizational knowledge: a study of the uses of coercive, mimetic and normative isomorphism", *Administrative Science Quarterly*, Vol. 44 No. 4, pp. 653-683.
- Moravesik, M. J. and Murugesan, P. (1975) "Some results on the function and quality of citations", *Social Studies of Science*, Vol. 5 No. 1, pp. 86-92.
- Nicolaisen, J. (2007) "Citation analysis", *Annual Review of Information Science and Technology*, Vol. 41, pp. 609-641.
- Oppenheim, C. and Renn, S. P. (1978) "Highly cited old papers and the reasons why they continue to be cited", *Journal of the American Society for Information Science*, Vol. 29 No. 5, pp. 225-231.
- Peritz, B. C. (1983) "A classification of citation roles for the social sciences and related fields", *Scientometrics*, Vol. 5 No. 5, pp. 303-312.
- Ritchie, A. 2008. *Citation Context Analysis for Information Retrieval*. PhD thesis, University of Cambridge.
- Rolling, L. (1981) "Indexing consistency, quality and efficiency", *Information Processing and Management*, Vol. 17 No. 2, pp. 69-76.
- Shadish, W. R., Tolliver, D., Gray, M. and Gupta, S. K. S. (1995) "Author judgements about works they cite: three studies from psychology journals", *Social Studies of Science*, Vol. 25 No. 3, pp. 477-498.
- Spiegel-Rosing, I. (1977) "Bibliometrics and content analysis", *Social Studies of Science*, Vol. 7 No. 1, pp. 97-113.
- Vinkler, P. (1987) "A quasi-quantitative citation model", *Scientometrics*, Vol. 12 No. 1/2, pp. 47-72.

- White, H. D. (2004) "Citation analysis and discourse analysis revisited", *Applied Linguistics*, Vol. 25 No. 1, pp. 89-116.
- Wilkinson, D., Harries, G., Thelwall, M. and Price, L. (2003) "Motivations for academic web site interlinkings: evidence for the Web as a novel source of information on informal scholarly communication", *Journal of Information Science*, Vol. 29 No. 1, pp. 49-56.

**Appendix I.** The ten articles from *Aslib Proceedings* Volume 63, issue 2/3 (2011) that were used in the citation study.

- Clough, P., Tang, J., Hall, M.M. and Warner, A. "Linking archival data to location: a case study at the UK National Archives", pp. 127-147.
- Chen, H., Baptista Nunes, M., Zhou, L. and Peng, G.C. "Expanding the concept of requirements traceability: The role of electronic records management in gathering evidence of crucial communications and negotiations", pp. 168 – 187.
- Idrees, I., Vasconcelos, A.C. and Cox, A.M. "The use of Grounded Theory in PhD research in knowledge management: A model four-stage research design", pp. 188 – 203.
- Cox, A.M. and Blake, M.K. "Information and food blogging as serious leisure", pp. 204 – 220.
- McKinney, P., Jones, M. and Turkington, S. "Information literacy through inquiry: A Level One psychology module at the University of Sheffield", pp. 221 – 240.
- Gumulak, S. and Webber, S. "Playing video games: learning and information literacy", pp. 241 – 255.
- Beverley, C.A., Bath, P.A. and Barber, R. "Health and social care information for visually-impaired people", pp. 256 – 274.
- Birdi, B. "Investigating fiction reader characteristics using personal construct theory", pp. 275 – 294.
- Corrall, S. and O'Brien, J. "Developing the legal information professional: A study of competency, education and training needs", pp. 295 – 320.
- Bakri, A. and Willett, P. "Computer science research in Malaysia: a bibliometric analysis", pp. 321 – 335.

## **Appendix II.** Classification of reasons for citing, based on work by Harwood (2009).

### Signposting

1. Helping and interesting less informed readers. This type of citation is used to provide introductory reading for those who may be unfamiliar with the subject of the current publication, e.g., a chapter in an edited book might mean that part of the audience will require citations to background material that would not have been required when writing for a specialist journal.
2. Keeping the argument on track. This type of citation is used to avoid tangential details, definitions, and methodological explanations, e.g., rather than describing some complex statistical test or large body of previous studies in detail, those interested can read the cited work.

### Supporting

3. Justifying the topic of research. This type of citation is used to show that the topic of the publication is of current or continuing interest, or to show that there is a gap or limitation in the published research that needs addressing
4. Justifying the method and methodology employed. This type of citation is used to support the particular approach used in a publication, e.g., if the author anticipates readers' and/or reviewers' objections to the handling of methodological issues.
5. Justifying claims. This type of citation is used to support the claims that are being made by the author in the current publication.

### Credit

6. Paying respect. This type of citation is used to reflect the debt that the author owes to previous publications, e.g., acknowledging an idea from an individual or a specific article that has provided the basis for the current publication.
7. Self-defence. This type of citation is used to acknowledge that others have already published in the area of the current publication, contrasting their particular approach with that in the current publication (possibly to avoid criticisms from reviewers)

### Position

8. Exemplars of positions. This type of citation is used to provide illustrative examples of different views that have been expressed in the literature, e.g. "A has said this whereas B suggests that...".
9. Detailed explication of positions and results. This type of citation is used to spell out the details of a previous publication in considerable detail, e.g., when proposing some subtle methodological improvement or when seeking to criticise previous work.
10. Tracing positions over time. This type of citation is used to summarize the origins and chronological development of a subject, or of a specific author's contribution to that subject.

### Engaging

11. Praising but then identifying problems with the source. This type of citation is used when the author is broadly in agreement with a previous publication but when the author wishes to note some flaw or limitation in the arguments presented therein.
12. Identifying inconsistencies in source's position. This type of citation is used when

there is a more marked disagreement with a previous publication, e.g., highlighting a major flaw or inconsistency, criticising the statistical methods used to analyse the data, or simply suggesting that the publication is seriously incorrect.

#### Building

13. Building by citing own work or that of others. Much academic research is necessarily incremental in nature, building on previous publications (and this is particularly the case with many self-citations): this type of citation is used so that readers can obtain the necessary basis for the work described in the current publication.

#### Tying

14. Tying in with others' methods. This type of citation is used to align the current publication with an existing methodological approach, e.g., to demonstrate that a standard, well-established approach is being used to analyse the data in a project.
15. Tying in with schools of thought. This type of citation is used to show that the current publication is aligned with a particular school of thought, methodological approach, or position in some ongoing debate within the discipline

#### Future

16. Mapping out future work planned by writer. This type of citation is used to identify sources that could form the basis for future work, e.g., "we intend to apply our methods to the dataset described by X"

#### Competence

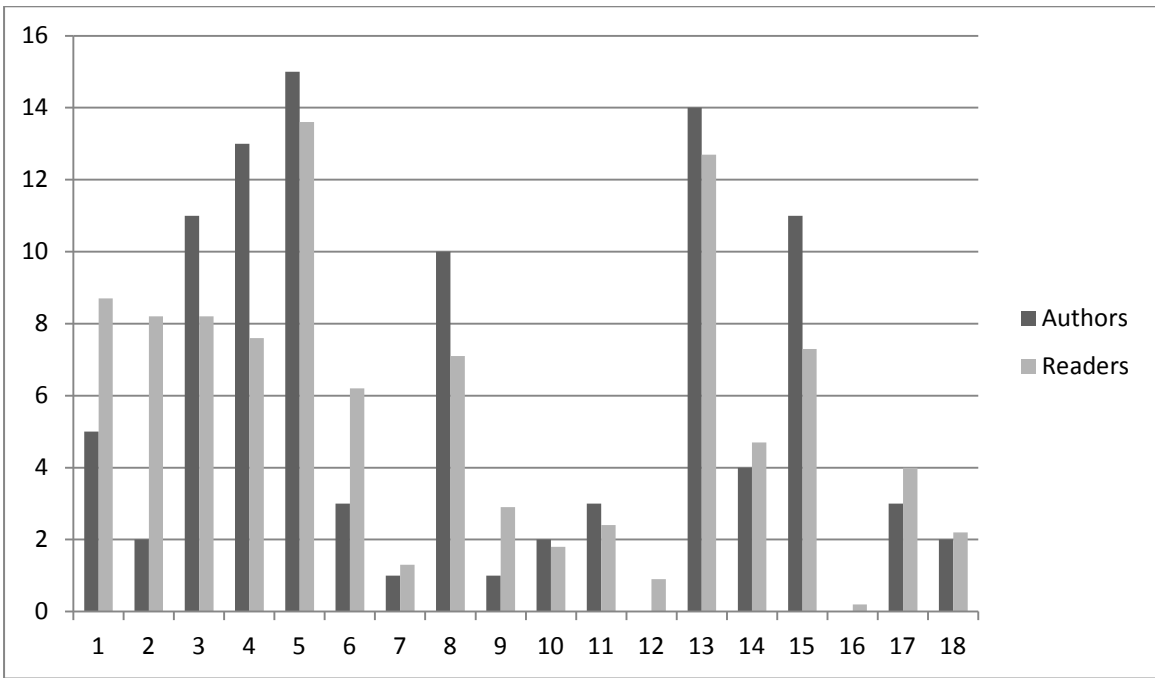
17. Displaying knowledge of the literature. This type of citation is used when citing key articles by leaders in the field, enabling authors to display their knowledge of the literature and thus underscoring their command and competence of the area.

#### Other

18. Other. Any type of citation that is not described by any of the 17 sub-classes above.

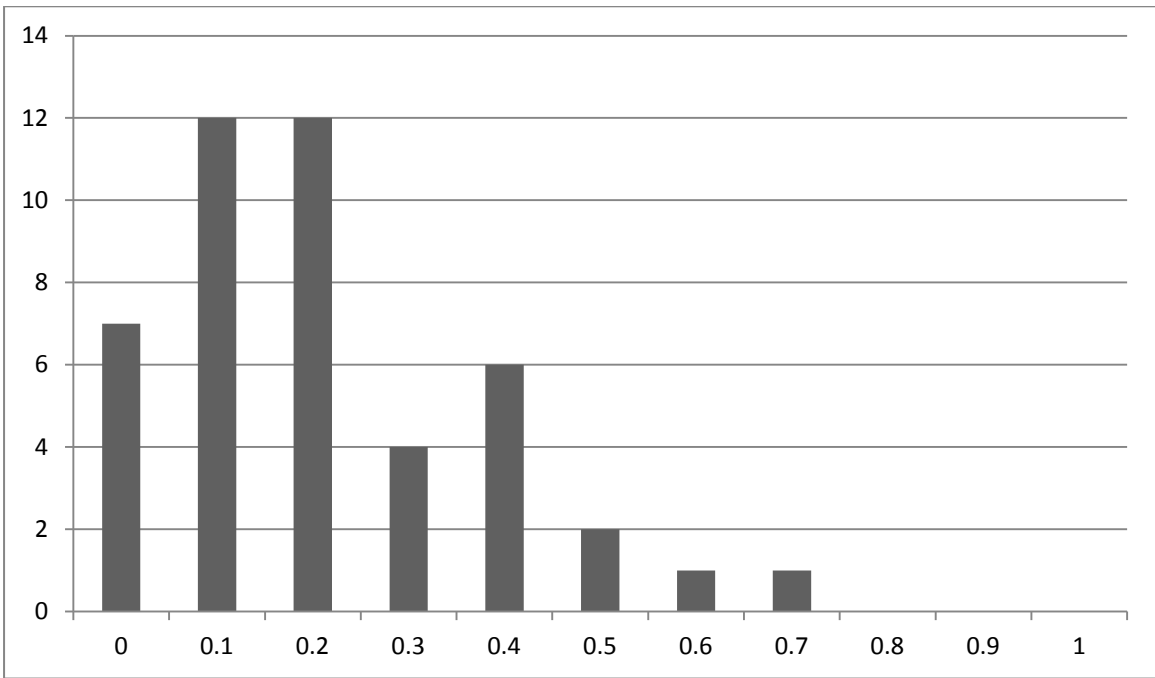
| Similarity | A  | B  | C  |
|------------|----|----|----|
| 0.0        | 7  | 4  | 0  |
| 0.1        | 12 | 6  | 2  |
| 0.2        | 12 | 11 | 6  |
| 0.3        | 4  | 10 | 10 |
| 0.4        | 6  | 5  | 8  |
| 0.5        | 2  | 5  | 6  |
| 0.6        | 1  | 3  | 7  |
| 0.7        | 1  | 1  | 3  |
| 0.8        | 0  | 0  | 2  |
| 0.9        | 0  | 0  | 0  |
| 1.0        | 0  | 0  | 1  |

**Table 1.** Distribution of author-reader Dice similarity values using: 18 possible reasons for citation (A); 10 possible reasons for citation (B); and 18 possible reasons including alternative suggestions (C). Each element in a row contains the number of author-reader comparisons that yielded that row's similarity value.



**Figure 1.** Bar-chart showing the distribution (as percentages of the total assignments) of the 18 possible reasons for citation assigned by the authors and by the readers.





**Figure 2.** Bar-chart showing the distribution of 45 author-reader Dice similarity values using 18 possible reasons for citation.