

# Are Wikipedia Citations Important Evidence of the Impact of Scholarly Articles and Books?<sup>1</sup>

Kayvan Kousha and Mike Thelwall

Statistical Cybermetrics Research Group, School of Mathematics and Computer Science, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, UK.

E-mail: {k.kousha, m.thelwall}@wlv.ac.uk

**Individual academics and research evaluators often need to assess the value of published research. Whilst citation counts are a recognised indicator of scholarly impact, alternative data is needed to provide evidence of other types of impact, including within education and wider society. Wikipedia is a logical choice for both of these because the role of a general encyclopaedia is to be an understandable repository of facts about a diverse array of topics and hence it may cite research to support its claims. To test whether Wikipedia could provide new evidence about the impact of scholarly research, this article counted citations to 302,328 articles and 18,735 monographs in English indexed by Scopus in the period 2005 to 2012. The results show that citations from Wikipedia to articles are too rare for most research evaluation purposes, with only 5% of articles being cited in all fields. In contrast, a third of monographs have at least one citation from Wikipedia, with the most in the arts and humanities. Hence, Wikipedia citations can provide extra impact evidence for academic monographs. Nevertheless, the results may be relatively easily manipulated and so Wikipedia is not recommended for evaluations affecting stakeholder interests.**

## Introduction

Citation-based indicators have been widely used in research evaluations to supply quantitative evidence of scholarly impact, typically to support qualitative judgements. Despite the many limitations of citation-based indicators for research evaluation (MacRoberts & MacRoberts, 1989, 1996; Moed, 2005), they can be used on a small scale for individual publications (Abramo & D'Angelo, 2015) or even on a large scale to compare the outputs of nations (Elsevier, 2013a). The citations used are typically derived from one of two major citation indexes, the Web of Science (WoS) and Scopus, which mainly index academic journal articles although they also include some conference proceedings and books (Elsevier, 2013b; Thomson Reuters, 2015). Free scholarly digital libraries, such as Google Scholar, may provide additional data when more comprehensive coverage is needed. Nevertheless, since these databases primarily index citations in the works of publishing scholars, they are not useful for direct evidence of societal, cultural or educational impacts. Hence, new sources of evidence are required in order to capture evidence of these wider impacts, and although a number of new indicators have been proposed for this, all are imperfect and new indicators are still needed (Wouters, et al., 2015).

Citations from encyclopaedia articles to academic publications may reflect the transfer of knowledge from the scholarly domain into a format that is accessible to, and perhaps used by, a wider public. According to Alexa.com's panel of toolbar users, in September 2015 Wikipedia was the seventh most

---

<sup>1</sup> This is a preprint of an article to be published in the Journal of the Association for Information Science and Technology © copyright 2015 John Wiley & Sons, Inc.

visited website in the world (<http://www.alexa.com/siteinfo/wikipedia.org>) and so it seems to be an important source of information for a section of the online public. Moreover, at least a third of American internet users consult Wikipedia and a majority of those with a college degree (Rainie & Tancer, 2007). Nevertheless, there have been concerns about the accuracy and completeness of Wikipedia articles (Chesney, 2006; Denning, Horning, Parnas, & Weinstein, 2005; Gorman, 2007; cf. Giles, 2005) because anonymous authors may add incorrect information (Mehegan, 2006). As a result, Wikipedia should be used cautiously in higher education (Chen, 2009; Chen, 2010; Eijkman, 2010; Luyt & Tan, 2010; Bayliss, 2013) and should not be used in academic publications (Cohen, 2007; Bould et al., 2014). Although Wikipedia is widely used for learning, teaching and other academic activities, (e.g., Aibar et al., 2015; Giles, 2005; Lim, 2009; Dooley, 2010; Head & Eisenberg, 2010; Knight & Pryke, 2012; Soules, 2015), it is not clear what the main audience for Wikipedia articles citing academic research is: students, academics or the general public. Nevertheless, it is at least possible that Wikipedia reflects knowledge production and use in a wider social context (Luyt & Tan, 2010) or has some “epistemic virtues” (e.g., power, speed, and fecundity) (Fallis, 2008) and hence citations from Wikipedia may indicate broader applications of academic outputs inside and/or outside of academia.

Despite many investigations into the reliability and accuracy of Wikipedia (Jullien, 2012; Mesgari et al. 2015), its readership among different groups (Okoli et al., 2014), its academic credibility (Nielsen, 2007; Luyt & Tan, 2010; Haigh, 2011; Park, 2011; Stankus & Spiegel, 2010), and its relationship to open access publishing (Teplitskiy, Lu, & Duede, 2015) few have analysed it as an alternative source of evidence about the broader impact of academic research (Priem, Piwowar, & Hemminger, 2012; Lin & Fenner, 2014) and there have been no in depth multidisciplinary evaluations so far. The current study partly fills this gap by investigating whether Wikipedia citations to scientific articles and books can be extracted automatically and if they are sufficiently numerous for the wider impact assessment of academic articles and books.

## Sources of Web Impact

There are many different ways to count citations from parts of the web. Early experiments searched for references to articles in web pages indexed by commercial search engines (Vaughan & Shaw, 2003, 2005). Later studies often focused on specific document types or websites, including preprints, dissertations, web CVs, presentation files, course reading lists, forums, news pages, and library websites (Thelwall & Kousha, 2015).

### *Scientific Impact*

Much academic and scholarly-related content is now on the web, including journal articles, conferences papers, patents and books. This has created the potential to extract information about scholarly impact from the web.

*Google Scholar:* Google Scholar provides greater overall coverage of publications and citations than traditional citation indexes such as WoS and Scopus (Meho & Yang 2007; Kousha & Thelwall, 2007a; Bar-Ilan, 2008). Moreover, retrospective coverage of Google Scholar has expanded recently, increasing its advantage (Harzing, 2014, de Winter, Zadpoor, & Dodou, 2014). It now seems to cover about 88% (100 million) of the English-language scholarly documents accessible on the web (Khabisa & Giles, 2014). Although Google Scholar lacks quality control procedures for selecting documents to index, in contrast to traditional citation indexes, and therefore it is possible to manipulate citation counts with fake documents (Jacso 2011; Labbé & Labbé, 2013; López-Cózar, Robinson-García, & Torres-Salinas, 2014), it

may still be essential for the impact assessment of scholarly publications in languages and countries that are not well covered by WoS or Scopus.

*Google Books:* Google Books indexes a large number of academic and other books and is an important source of scientific citations from books (Kousha & Thelwall, 2009). Although it is not a citation index, citations can be extracted from the Google Books API with appropriate queries (Kousha & Thelwall, 2014). Google Books citations are more numerous than Scopus citations (Kousha, Thelwall, & Rezaie, 2011) and the Thomson Reuters' Book Citation Index citations (Kousha & Thelwall, 2014) in the humanities and some social sciences, and so it is particularly useful impact indicator for book-based disciplines. Although most volumes within the Google Books database probably do not have academic authors, those that cite scholarly research seem likely to have authors from within academia. Hence, counting citations to academic research from Google Books seems likely to predominantly reflect a type of scientific impact.

### *Wider Impacts*

The web may reflect impacts of research outside of the publishing scholarly community through the many different types of document that are posted online.

*Google Patents:* Scientific references in patents may reflect the commercial or technical value of academic publications (Narin & Olivastro, 1992; Schmoch, 1993). Google Patents allows full-text searching of patents from the United States Patent and Trademark Office (USPTO), the European Patent Office (EPO) and other patent offices. Similarly to Google Books, it is possible to identify citations to scientific publications in the non-patent references. A semi-automatic method can extract citations from Google-indexed patents on a large scale and low correlations between patent citations and journal citations suggest that patent citations may reflect the wider commercial value of research (Kousha & Thelwall, in press).

*Presentations:* Citations from online presentations (e.g., in PowerPoint format) may reflect the early scholarly impact of research from conference presentations of early work, including in conference-based fields (Thelwall & Kousha, 2008). They may also indicate educational uptake in the case of presentations delivered as part of academic courses or student talks.

*Syllabus mentions:* Mentions of publications in academic course syllabi are evidence of their value for teaching, as judged by instructors (Kousha & Thelwall, 2008). Syllabus mentions are a useful indicator for the educational impacts of books and textbooks. This is especially important in the social sciences and the humanities, where academic outputs may have educational value rather than, or in addition to, research impact. Weak but statistically significant correlations between academic syllabus mentions and traditional citations in many fields are broadly consistent with this conclusion (Kousha & Thelwall, 2015).

*Science blogs:* Citations in science blog posts (e.g., ResearchBlogging.org and ScienceBlogs.com) suggest the transmission of ideas to a wider educated public as well as other scholars when bloggers reference scientific publications to discuss, disseminate or publicise research in an informal way (Kovic, Lulic, & Brumini, 2008; Luzón, 2009). Science blog citations have also been found to have a positive but low correlation with citation counts (Adie & Roe, 2013; Shema, Bar-Ilan, & Thelwall, 2014).

*Clinical guidelines:* Online clinical guidelines could be useful to assess the societal impact of individual medical studies by demonstrating their worth in deciding upon general health care solutions (Kryl, Allen, Dolby, Sherbon, & Viney, 2012). A study of the cited references of clinical guidelines produced by the National Institute of Health and Clinical Excellence (NICE) in the UK found that articles cited in guidelines tend to be more highly cited than comparable articles (Thelwall & Maflahi, 2015).

# Wikipedia

## *Uses of Wikipedia*

The use of an encyclopaedia is predicated on a perception of credibility, especially in an academic context (Mesgari et al., 2015). Wikipedia recommends that “articles should be based on reliable, third-party, published sources with a reputation for fact-checking and accuracy” ([https://en.wikipedia.org/wiki/Wikipedia:Identifying\\_reliable\\_sources](https://en.wikipedia.org/wiki/Wikipedia:Identifying_reliable_sources)). Although opinions about the credibility of Wikipedia are mixed, most seem to accept that it has some merits. A comparative analysis of book citations in 47 entries about the brain or behavioural sciences in Wikipedia and Scholarpedia (a peer-reviewed open-access encyclopaedia) found that both encyclopaedias included references to reputable information (Stankus & Spiegel, 2010). Nevertheless, a survey of academics around the world (n=201) has shown that there are concerns about using Wikipedia (Chen, 2010) and another survey of university faculty members in the U.S. (n=105) found that none ranked Wikipedia as “extremely credible”. Only 3% of academics ranked Wikipedia as “very credible” compared with about 20% who declared that Wikipedia has “no credibility” (Dooley, 2010). Individual topics can also be controversial (e.g., Yasseri, Spoerri, Graham, & Kertész, 2014) and so it would be difficult to claim objectivity in all cases.

Wikipedia is accessed in universities by both academics and students. A survey of 1,000 Nature authors found that 17% consulted Wikipedia weekly but less than 10% helped to update it (Giles, 2005). A third of U.S. college students in another survey (n=134) used Wikipedia for academic purposes, and especially for background information related to their courses (Lim, 2009). Similarly, from 2,318 college students at six different U.S. colleges, most either always (30%) or frequently (22%) used Wikipedia to obtain course-related background information (Head & Eisenberg, 2010). A survey of 137 academic staff from four universities in Australia and one each in Canada, the UK and South Africa found that over a half (56%) accepted student use of Wikipedia as a “research starter”, although 22% either discouraged or strongly disapproved of its use (Eijkman, 2010). At Liverpool Hope University, 75% of academics (out of n=133) and students (out of n=1,222) surveyed used Wikipedia for academic teaching and learning (Knight & Pryke, 2012). About half of 913 faculty members in two public universities in Spain (47%) agreed that Wikipedia was a useful teaching source, whereas only 19% disagreed. Moreover, 27% claimed that they frequently or very frequently recommended students to consult Wikipedia (Aibar et al., 2015). A quarter of university faculty members at a university in the U.S. (n=105) reported that they “don’t ever use Wikipedia in teaching/research” and 40% declared that they occasionally use Wikipedia in academic activities (Dooley, 2010). Hence, Wikipedia is extensively, but not universally, used in higher education and by researchers.

Outside of academia, Wikipedia is a prominent source of general medical and healthcare information (e.g., Kim et al., 2010; Devgan, Powe, Blakey, & Makary, 2007; Heilman et al., 2011; Thomas, Eng, de Wolff, & Grover, 2013). Over 155,000 medical Wikipedia articles in different languages were accessed more than 4.88 billion times in 2013, making it one of the most viewed medical and health care resources on the internet (Heilman & West, 2015). It is also frequently listed in search engine results for health-related queries (Laurent & Vickers, 2009). Wikipedia is a valuable resource for medical professionals: about 70% of junior physicians from a major London medical school (n=35) used it for their clinical decisions and medical education (Hughes, Joshi, Lemonde, & Wareham, 2009) and a usage-log analysis of medical students in Australia (n=842) found that Google (70%) and Wikipedia (51%) were

more used than eMedicine (21%), the National Institutes of Health (NIH) (16%) and the university library (Judd & Kennedy, 2011).

Wikipedia is consulted by the public for a variety of information needs, many of which will not need current academic research to help satisfy. A study of the 100 most-visited Wikipedia pages (September 2006 to January 2007) showed that entertainment (43%), politics and history (15%), geography (12%), and sexuality (10%) were the most popular, whereas only 6% were on scientific topics (Spoerri, 2007). It is not clear whether the non-scientific topics often cite academic research, however. Presumably this would be most likely for politics and history or geography and least likely for entertainment.

### *Citations to Wikipedia Articles*

Evidence of the relationship between Wikipedia and scholarship can be found in the Wikipedia articles that are cited in scientific publications. A study of English medical science journals indexed in Medline, PubMed, or Embase found that 1,433 articles from 1,008 journals cited 2,049 Wikipedia articles. The majority of the citations to Wikipedia articles were created for definitions (31.6%), descriptions of processes (23.5%), or historical background (13.5%), whereas less than 5% cited it instead of the original research discussed in the cited article (Bould et al., 2014). Similarly, articles in chemistry journals from three major publishers (the American Chemical Society, Elsevier, and Springer) in the period of 2005 to 2009 rarely (370 times in total) cited Wikipedia and most Wikipedia articles (63%) were cited for “general scientific information”, typically relating to science, technology, engineering, and mathematics (Brazzeal, 2011)

The number of Scopus publications citing Wikipedia has increased dramatically over time, however (Figure 1). A Scopus cited reference search in August 2015 identified 61,136 Scopus publications citing at least one Wikipedia article compared with 7,849 Scopus publications citing at least one Encyclopaedia Britannica article. While there were 0 citations to Wikipedia articles from Scopus publications in 2001 and 12 in 2003, this had increased to 8,579 by 2014 (Authors’ own data: <http://dx.doi.org/10.6084/m9.figshare.1512806>). Nevertheless, if the result is normalised by the total number of Wikipedia articles (e.g., 4.9 million English articles: [https://en.wikipedia.org/wiki/Wikipedia:Size\\_in\\_volumes](https://en.wikipedia.org/wiki/Wikipedia:Size_in_volumes)), then these numbers are relatively low.

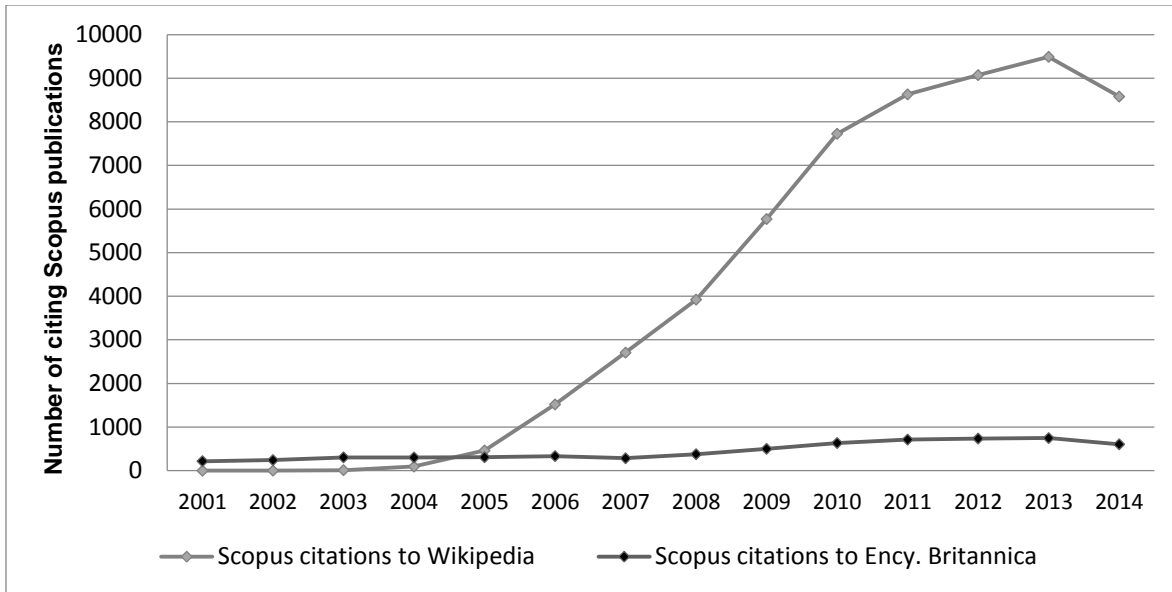


Figure 1. Scopus-indexed citations to Wikipedia and Encyclopaedia Britannica articles (2001-2014), as of August 2015. Citations to Encyclopaedia Britannica articles were estimated using cited references source title search (REFSRCTITLE) for “\*Encyclopædia Britannica\*”, “\*Encyclopedia Britannica\*”, “\*Encyclopaedia Britannica\*”, “\*Britannica Online Encyclopaedia\*” or “\*Britannica Online Encyclopedia\*” in combination of cited reference search (REF) for “\*britannica.com\*” to capture citations to online Encyclopaedia Britannica articles. The query REF(“\*wikipedia.org/wiki\*”) was used for Wikipedia.

### Wikipedia-Cited Scientific Publications

Several investigations have assessed references in Wikipedia articles. A study of ISBN, PubMed, DOI and ArXiv identifiers in English Wikipedia showed that the majority of matches were to books and monographs (Halfaker & Taraborelli, 2015). Over a third (35%) of the cited references were to books compared with less than 2% to academic journal articles from a sample of 50 national history articles (Luyt & Tan, 2010). About 62% of the references used in the Wikipedia articles were internet sources, such as government websites and news and media, suggesting that Wikipedia is “a product of a wider social context” rather than a “serious reference work” (Luyt & Tan, 2010, p. 721). These studies also suggested that books and monographs are more frequently cited in Wikipedia entries than are academic articles.

Relatively few articles ever get cited in Wikipedia. Only 4-5% of articles published by the Public Library of Science (PLOS) have been cited in Wikipedia (Lin & Fenner, 2014; Priem, Piwowar, & Hemminger, 2012). One study found a significant correlation between counts of citations from Wikipedia to articles in scientific journals and their impact factors as reported by Journal Citation Reports, suggesting that Wikipedia articles tend to cite articles in high impact journals such as *Nature*, *Science* and the *New England Journal of Medicine* (Nielsen, 2007). An analysis of characteristics of publications cited in Wikipedia found that the most cited publications were from information science and computer science, however, and that the *Journal of American Society for Information Science and Technology* and *Lecture Notes in Computer Science* were amongst the most popular (Park, 2011).

In the field of medical science, references in medical-related Wikipedia articles increased by more than 2.5 times from 2009 to 2013 across all languages. *The Lancet*, *The New England Journal of*

*Medicine, Nature, British Medical Journal, The Journal of the American Medical Association (JAMA) and Science* were the most cited medical journals in medical Wikipedia pages (Heilman & West, 2015). Similarly, an investigation on sample of 50 health related Wikipedia entries (e.g., Cancer, Alzheimer, smoking) showed that 56% of the cited references were from reputable sources such as *the New England Journal of Medicine, Nature, JAMA, and Archives of Internal Medicine* (Haigh, 2011).

In summary, books seem to be cited more often than articles, reputable journals are often cited, and health, information science and computer science are common topics for citations from Wikipedia.

## Research Questions

The aim of this paper is to analyse the potential for Wikipedia citations to be used as a new source of evidence about the educational, societal, cultural, or other impacts of articles and books that are not well reflected by their citation counts. For example: a course textbook might be cited in relevant Wikipedia articles as a result of students finding it useful; a monograph might be cited in an article on local government policies because it is widely used by local government officials, at least one of whom edited Wikipedia; a popular work on renaissance art might be cited in several relevant artists' biographies on Wikipedia by interested art connoisseurs that appreciate the historical insights that it has given them during recreational reading. These types of impacts are important for evaluators to assess, especially in the arts, humanities and social sciences, where cultural or educational outcomes can be important. As a first step, it is logical to assess whether they are frequent enough to be worth counting and the extent to which they correlate with traditional citation counts across fields and years (Sud & Thelwall, 2014). A perfect correlation (+1) would render Wikipedia citations redundant, whereas other statistically significant positive correlation would be evidence that Wikipedia citations were not random and could, together with other evidence, support an argument for how they should be interpreted.

1. Are academic articles and monographs cited often enough in Wikipedia for it to be a useful source of impact evidence?
2. Do Wikipedia citations reflect impacts of academic research outside of academia (e.g., cultural, societal and educational)?
3. Are there disciplinary differences in the answers to the above questions and between papers and monographs?

Although, as reviewed above, there have been partial answers to the above questions in previously-published research, such as for individual journals or topic areas, a large scale systematic study across multiple fields, years and academic outputs (e.g., books) is needed in order to give general answers and conclusions about the implications for research evaluation. Moreover, comparisons with other alternative indicators, such as citations from monographs and academic syllabus mentions, are also needed to give insights into whether the educational or cultural impacts of research are reflected in Wikipedia citations.

## Methods

The research design is to create large samples of recent academic journal articles and books from multiple disciplines and to assess the extent to which they are cited in Wikipedia through comparisons with traditional and other types of citation counts. These other types are academic syllabus mentions, which reflect educational impact, and Google Books citations, which represent book-based impact. Both of these are relevant to Wikipedia, as confirmed by the above literature review. Spearman correlations

between the different counts were used to assess, in part, the extent to which they reflect similar types of impact.

### *Data Sets*

The investigation was based on counts of Wikipedia citations to 302,328 articles and 18,735 monographs from multiple Scopus subject areas published during 2005-2012. The Scopus citation database was used instead of the Thomson Reuters Web of Science because it has better coverage of arts and humanities academic journals and books (Meester, 2013) and it is adequately mature, especially from 1996 onwards (Ball & Tunger, 2006). About 21,000 peer-reviewed journals (Scopus content coverage guide, 2014) and 117,000 academic books (Authors' data: <http://dx.doi.org/10.6084/m9.figshare.1584632>) were indexed by Scopus at the time of the study compared with about 12,000 scientific journals (Web of Science core collection, 2015) and 60,000 books (The Book Citation Index in Web of Science) in the Web of Science.

These years were selected to give a long enough time window (eight years) for comparisons over time whilst giving academic articles and books at least two years to attract citations. Comparisons over time are useful in case the practice of Wikipedia citation is evolving rapidly. For Scopus articles, bibliographic information and citation counts for English language articles and reviews (omitting editorial materials, letters, notes, book chapters, short surveys and erratum) were extracted from sixteen science, medicine, social science and humanities fields in order to give a wide selection of different academic areas. Subject areas were selected based on Scopus All Science Journal Classification (ASJC) codes (see: [http://ebrp.elsevier.com/pdf/Scopus\\_Custom\\_Data\\_Documentation\\_v4.pdf](http://ebrp.elsevier.com/pdf/Scopus_Custom_Data_Documentation_v4.pdf)). The advanced Scopus search command 'subjmain' (e.g., subjmain(1202) for History) was used to limit the results to a specific field. Very distinct subject categories were then removed from the Scopus results (e.g., 'Dentistry' or 'Chemistry' from History search results) to give a more focused dataset for comparisons between disciplines. However, we included results within subjects that were more homogeneous and with a large overlap between Scopus categories (e.g., Social Sciences and Arts and Humanities disciplines).

For each selected field and year, a random sample of 2,500 articles was taken from all downloaded Scopus records to have large enough data sets for analyses in different years (2,500 \* 8 years = 20,000). Records without author information in Scopus were removed in order to perform effective searches. Articles with less than three words in their titles were excluded due to many records being incorrectly indexed as articles in Scopus (e.g., Editorial, Commentary, Conclusion, Forward, Editor's note, This issue) and articles with very short titles (e.g., Hero, Fever, Tokyo) that could generate many false matches in the Wikipedia searches used. Articles with one or two words in their titles were more common in Music, Visual Arts and Performing Arts, Philosophy, and Literature than in other fields.

In order to compare Wikipedia citations to monographs with Wikipedia citations to articles, a random sample of 500 Scopus books was taken in the Arts and Humanities and Social Sciences in each year (500 \* 8 years = 4,000). For the other fields analysed with fewer indexed books, all Scopus-indexed volumes were used. The Scopus advanced search command "*DOCTYPE (bk)*" was used and the results were restricted to English books published during 2005-2012 in order to have a homogeneous set. Edited books, book series, book chapters, trade publications and books without author information ("[No author name available]") were excluded in order to focus on monographs. To avoid multiple book editions, records with information in Volumes or Editors fields and book titles including terms such as



“edition” (e.g., second edition) or “volume” (e.g., Vol. III) were also excluded. This decision was made because Scopus citations for individual editions are not included within the counts of citations to other editions and hence citation counts could be underestimated for edited volumes. For instance, the tenth edition of book *“Introduction to Probability Models”* by S.M Ross published in 2009 had no Scopus citations, whereas its eleventh edition published in 2014 had received 8 Scopus citations at the time of study. This problem has already been identified for bibliometric analyses of edited books and volume series using the Thomson Reuters’ Book Citation Index (Leydesdorff & Felt, 2012). Due to problems in Scopus with limiting book results to a specific subject area, especially in the humanities (e.g., history or education), the book searches were restricted to the eleven broad subject areas. Multidisciplinary books were also ignored by removing multiply-classified books (e.g., excluding “Arts and Humanities” monographs from “Medicine” search matches). Similarly to the case for articles (see above), monographs with titles containing less than three words were excluded to avoid retrieving many false matches from the Wikipedia searches. For both Scopus-indexed articles and monographs, Scopus citation counts (the “Cited by” field) were used.

### *Automatic Wikipedia Citation Searches*

To identify citations to books or articles from Wikipedia, the simplifying assumptions were made that such citations would not include typos and would include some basic information about the cited work, such as the author and title, and that citations could be identified through search engine queries. As a result of these assumptions, partial coverage of the web by search engines and incomplete sets of search engine results (Bar-Ilan, 2004), the Wikipedia citation counts described below are likely to be underestimates.

The Bing API in the free software Webometric Analyst (<http://lexiurl.wlv.ac.uk>) was used to automatically search for citations in Wikipedia. The queries were automatically generated from the Scopus bibliographic information by Webometric Analyst (see the “Make Wikipedia Searches for Scopus/WoS Data” option in the “Make Searches” menu) which was designed for this purpose. To generate effective queries the first (up to) three authors’ last names were used, the first (up to) ten terms of the article title as a phrase search, and the publication year, together with the *site:wikipedia.org/wiki/* command to limit the results to Wikipedia articles, as shown in the examples below.

*Choi "Obesity weight change hypertension diuretic use and risk of gout" 2005 site:wikipedia.org/wiki*

*Rupp Breunig Ameje "New studies on the Nok culture of Central Nigeria" 2005 site:wikipedia.org/wiki*

For articles with three or four words in their titles, journal names were also added to reduce the number of false matches from the searches. This reduces the number of false matches but may also lose some correct matches that used an abbreviated journal name (see examples below).

*Weder "Genetics and hypertension" "Journal of Clinical Hypertension" 2007 site:wikipedia.org/wiki*

*Elliott Gibbs "Does dyslexia exist" "Journal of Philosophy of Education" 2008 site:wikipedia.org/wiki*

To locate citations to monographs in Wikipedia, publisher names were used instead of journals names for books with three to six terms in their title. This technique was used because books titles tend to be shorter and more general than scientific article titles and this could cause false matches. Publisher names in Scopus can be mentioned in multiple different ways (e.g., Springer US, Springer New York,

Springer Netherlands, Elsevier Ltd., Elsevier Inc.). Hence, the publishers' names were manually checked and standardised before use to give the most matches. For instance, "John Wiley & Sons" was used instead of "John Wiley & Sons Publishing", "John Wiley & Sons, Inc.", "John Wiley & Sons, Ltd." and "John Wiley and Sons" (see the examples below).

*Palsson "Systems biology Properties of reconstructed networks" "Cambridge University Press" 2006  
site:wikipedia.org/wiki/*

*Neiberg "The second battle of the Marne" "Indiana University Press" 2008 site:wikipedia.org/wiki/*

### *Google Books Citations*

Google Books seems to be the best available source of Book-based impact evidence. For the 18,735 monographs in the study, Google Books API citation searches in Webometric Analyst ("Books" tab) were conducted. These automatically extract citations from digitised books indexed by Google Books and remove false matches (for method details see: Kousha & Thelwall, 2014). Google books citations were used because Google books citations are more numerous than Scopus citations – at least for 1,000 books in the 2008 UK RAE in seven fields (Kousha, Thelwall, & Rezaie, 2011), and may reflect a different type of impact to that of journal articles. The Wikipedia citation queries were used to locate citations from Google Books, but omitting the command `site:wikipedia.org/wiki/` at the end of each query, as in the examples below.

*Henham Behrens "The criminal law of genocide International comparative and contextual aspects" 2008*

*Smith Vromen Cook "Keywords in Australian politics" "Cambridge University Press" 2006*

### *Academic Syllabus Mentions*

Mentions of books in course reading lists can reflect their teaching value (Kousha & Thelwall, 2008) and hence citations from online academic course syllabi were counted in order to investigate the teaching value of articles cited in Wikipedia (for method details see: Kousha & Thelwall, 2015). Book titles, authors and years were combined with either "syllabus" or "course description" to construct the queries. Bing API searches were used together with a set of results filtering rules to identify syllabus mentions accessible in the open web in over 25,000 academic websites and to exclude false matches (again in Webometric Analyst). The vertical bar "|" between queries in the example below is an OR operator to run two queries at the same time with the program combining the results after removing duplicates.

*Norris "Sacred and secular Religion and politics worldwide" 2011 syllabus|Norris "Sacred and secular Religion and politics worldwide" 2011 "course description"*

## Results

The results are organised to align with the first two research questions, with the third research question (disciplinary differences) being discussed within these two sections, when relevant.

### ***The prevalence of Wikipedia citations***

Addressing the first research question by assessing the prevalence of Wikipedia citations, the number of Scopus citations (3,684,506) to the sample of articles is significantly higher than the number of Wikipedia citations (36,191) to them (Table 1). About 76% out of the 302,328 articles had one or more Scopus citations, whereas only 5% had at least one Wikipedia citation, suggesting that relatively few

scientific papers are cited in Wikipedia entries. Hence, it seems that Wikipedia citations are not common enough to be used for the impact assessment of articles in most fields. Nevertheless, in History (11%) and in Music, Visual Arts and Performing Arts, Political Science and Astronomy and Astrophysics (7%) more articles in each field had at least one Wikipedia citation, perhaps reflecting popular interest in these areas, such as for biographies, history and popular science. Only 0.7% of the articles (2,133 of 302,328) had at least one Wikipedia citation but no Scopus citations, and so very few articles are recognised by Wikipedia alone. In contrast, 72% (218,169 out of 302,328) of articles had at least one Scopus citation but no Wikipedia citations. Of the articles with one or more Wikipedia citations but no Scopus citations, 71% (1,518 of 2,133) are from the arts and humanities (e.g., Music, Visual Arts and Performing Arts, History and Literature) and 20% are from the social sciences (e.g., Political Science and Education and Business), in comparison to the low figures for science (7%) and medicine (2%). For instance, the most Wikipedia-cited articles is “*Forgotten faces: Why some of our cinema heritage is part of the public domain*” by David Pierce, published in *Film History: An International Journal* with 85 Wikipedia citations and no Scopus citations. Most (17) of the 20 articles with the most Wikipedia citations but no Scopus citations are also from the arts and humanities, including many with cultural themes, such as films, media, music, literature and cultural studies. This suggests that Wikipedia citations may reflect the cultural value of academic outputs (Table 8 in the appendix) and types of arts and humanities impacts (e.g., educational or cultural) that are not well reflected in journal citations.

**Table 1.** Wikipedia citations extracted by the Bing API searches compared to Scopus citations for articles published during 2005-2012 (n=302,328).

<b>Scopus category</b>	Sample articles	Wikipedia citations No. (% with citations) median (mean) max	Scopus citations No. (% with citations) median (mean) max
Astronomy and Astrophysics	19,889	7,629, <b>(7.1%)</b> 0(0.4) 334	339,743, <b>(89.2%)</b> 8(17.1) 3690
Biochemistry, Genetics and Molecular Biology	19,750	2,857, <b>(6.2%)</b> 0(0.1) 25	513,996, <b>(85.9%)</b> 8(26) 7724
Computer Science	19,864	493, <b>(1.4%)</b> 0(0.02) 24	100,454, <b>(55.2%)</b> 1(5.1) 2069
Environmental Science	19,927	1,834, <b>(3.4%)</b> 0(0.1) 92	361,913, <b>(88.6%)</b> 9(18.2) 733
Internal Medicine	18,752	1,425, <b>(3.7%)</b> 0(0.1) 32	472,996, <b>(91.1%)</b> 11(25.2) 3005
Microbiology (Medicine)	19,864	2,031, <b>(4.3%)</b> 0(0.1) 72	360,034, <b>(93.9%)</b> 10(18.1) 1422
Pharmacology and Pharmaceutics	19,659	1,871, <b>(4.4%)</b> 0(0.1) 34	272,226, <b>(86.6%)</b> 7(13.8) 495
Surgery	18,280	542, <b>(2%)</b> 0(0.03) 23	210,326, <b>(86.1%)</b> 5(11.5) 392
Business, Management and Accounting	19,267	1,264, <b>(3.6%)</b> 0(0.1) 31	240,811, <b>(79.4%)</b> 4(12.5) 879
Education	19,693	842, <b>(2.7%)</b> 0(0.04) 20	168,521, <b>(79.7%)</b> 3(8.6) 1418
Psychology (clinical)	19,727	2,337, <b>(5.7%)</b> 0(0.1) 55	366,522, <b>(92.1%)</b> 9(18.6) 818

Political Science and International Relations	18,550	2,781, <b>(6.8%)</b> 0(0.1) 73	109,012, <b>(71.4%)</b> 2(5.9) 456
History	19,489	4,929, <b>(10.7%)</b> 0(0.3) 78	66,965, <b>(62%)</b> 1(3.4) 299
Literature	17,513	1,588, <b>(5.2%)</b> 0(0.09) 19	20,494, <b>(41.8%)</b> 0(1.2) 112
Music, Visual Arts and Performing Arts	15,789	2,314, <b>(7.1%)</b> 0(0.1) 82	22,806, <b>(43.2%)</b> 0(1.4) 110
Philosophy	16,315	1,454, <b>(5.3%)</b> 0(0.1) 22	57,687, <b>(64.4%)</b> 1(3.5) 165
Total	302,328	36,191, <b>(5%)</b> 0(0.1) 334	3,684,506, <b>(76.4%)</b> 4(12.2) 7724

\* Significant at the p = 0.05 level. \*\* Significant at the p = 0.01 level.

A third (33%) of the 18,735 Scopus monographs in all fields had at least one Wikipedia citation, whereas about 29% had one or more Scopus citations (Table 2). In the Arts and Humanities, just under of half (48%) and in the Social Sciences 39% of the monographs had at least one Wikipedia citation, which is almost double the rates for Scopus citations (26% and 16% respectively). Hence, Wikipedia citations are particularly plentiful in book-based fields. In Immunology and Microbiology (34%) and Psychology (32%) about a third of the Scopus books had at least one Wikipedia citation. However, in Medicine, Biochemistry, Genetics and Molecular Biology, Business, Computer Science, and Pharmacology and Pharmaceutics about a quarter (ranging from 22%-26%) of the monographs had at least one Wikipedia citation. In all areas Wikipedia cites a much higher percentage of books than journal articles, however, and a substantial minority of the Scopus-indexed books tested, and so it may be useful for book evaluations in all fields. Books are up to twice as likely to be cited by other books as they are to be cited by Wikipedia, although the difference varies by field (Table 2). Hence, Wikipedia citations should not replace citations from books.

**Table 2.** Wikipedia citations extracted by the Bing API searches and Google Books and Scopus citations for monographs published during 2005-2012 (n=18,735).

Scopus category	Monographs	Wikipedia Citations No. (% with citations) median (mean) max	Google Books Citations No. (% with citations) median (mean) max	Scopus Citations No. (% with citations) median (mean) max	Syllabus mentions No. (% with citations) median (mean) max
Arts and Humanities	3,762	10,839, <b>(48%)</b> 0(2.9) 603	31,382, <b>(77.5%)</b> 4(8.3) 62	25,902, <b>(25.9%)</b> 0(6.9) 791	3,983, <b>(32.2%)</b> 0(1.1) 70
Social Sciences	3,807	5,626, <b>(38.9%)</b> 0(1.5) 60	33,140, <b>(72.8%)</b> 4(8.7) 56	16,908, <b>(16.1%)</b> 0(4.4) 488	5696, <b>(35.7%)</b> 0(1.5) 123
Psychology	1,680	1,488, <b>(31.9%)</b> 0(0.9) 32	9,932, <b>(64.9%)</b> 2(5.9) 47	24,959, <b>(39.5%)</b> 0(14.9) 2590	1483, <b>(31.0%)</b> 0(0.9) 43
Business	1,099	519, <b>(23.4%)</b> 0(0.5) 29	6,089, <b>(65.2%)</b> 2(5.5) 38	13,937, <b>(56.2%)</b> 2(12.7) 372	877, <b>(25.5%)</b> 0(0.8) 36
Biochemistry, Genetics and Molecular Biology	1,351	953, <b>(23.2%)</b> 0(0.7) 38	2,371, <b>(39.7%)</b> 0(1.8) 39	13,294, <b>(44.6%)</b> 0(9.8) 913	488, <b>(16.2%)</b> 0(0.4) 21

Computer Science	1,572	1,231, <b>(25.6%)</b> 0(0.8) 46	4,175, <b>(43.6%)</b> 0(2.7) 42	15,251, <b>(35%)</b> 0(9.7) 856	1720, <b>(30.5%)</b> 0(1.1) 48
Environmental Science	1,472	1,589, <b>(28.4%)</b> 0(1.1) 142	5,702, <b>(55.3%)</b> 1(3.9) 46	6,140, <b>(24.2%)</b> 0(4.2) 196	705, <b>(18.3%)</b> 0(0.5) 21
Physics and Astronomy	1,132	1,885, <b>(31.4%)</b> 0(1.7) 70	1,562, <b>(29.0%)</b> 0(1.4) 37	33,121, <b>(30.4%)</b> 0(29.3) 4942	578, <b>(16.3%)</b> 0(0.5) 60
Medicine	2,400	1,692, <b>(21.8%)</b> 0(0.7) 48	5,375, <b>(43.7%)</b> 0(2.2) 39	12,082, <b>(16%)</b> 0(5) 3107	944, <b>(11.8%)</b> 0(0.4) 54
Immunology and Microbiology	161	177, <b>(34.2%)</b> 0(1.1) 34	284, <b>(36.0%)</b> 0(1.8) 46	423, <b>(32.9%)</b> 0(2.6) 74	22, <b>(8.1%)</b> 0(0.1) 4
Pharmacology and Pharmaceutics	299	474, <b>(24.7%)</b> 0(1.6) 191	565, <b>(52.2%)</b> 1(1.9) 30	3,915, <b>(65.6%)</b> 2(13.1) 518	59, <b>(12.4%)</b> 0(0.2) 5
All Fields	18,735	26,473, <b>(33.2%)</b> 0(1.4) 603	110,469, <b>(62.3%)</b> 2(5.9) 62	165,932, <b>(28.6%)</b> 0(8.9) 4942	16,561, <b>(25.9%)</b> 0(0.9) 123

Although there is a tendency for older books to be more cited by Wikipedia, other books and articles (Table 3), three years (i.e., the 2012 data) is long enough for half of the books that will eventually be cited by Wikipedia (taking the 2015 figure as approximately final). In terms of the length of time window required for half of all Scopus-indexed academic monographs to receive Wikipedia citations, the results suggests that in Arts and Humanities the minimum time would be five years and ten years for the Social Sciences.

**Table 3.** Monographs with at least one Wikipedia, Google Books, or Scopus citation by year and broad field (the anomalous figure of 29% for Arts and Humanities in 2007 has been double checked).

Broad fields	<b>% of monographs with one or more Wikipedia citations (in bold)</b>							
	<i>% of monographs with one or more Google Books citations (in italic)</i>							
	% of monographs with one or more Scopus citations							
	2005	2006	2007	2008	2009	2010	2011	2012
Arts and Humanities	<b>62%</b>	<b>60%</b>	<b>29%</b>	<b>54%</b>	<b>54%</b>	<b>49%</b>	<b>40%</b>	<b>37%</b>
	<i>90%</i>	<i>86%</i>	<i>79%</i>	<i>86%</i>	<i>74%</i>	<i>69%</i>	<i>75%</i>	<i>61%</i>
	19%	21%	30%	27%	27%	27%	35%	22%
Social Sciences	<b>62%</b>	<b>43%</b>	<b>40%</b>	<b>39%</b>	<b>33%</b>	<b>35%</b>	<b>26%</b>	<b>26%</b>
	<i>90%</i>	<i>84%</i>	<i>81%</i>	<i>79%</i>	<i>77%</i>	<i>73%</i>	<i>62%</i>	<i>58%</i>
	19%	20%	26%	28%	35%	34%	31%	32%
Sciences	<b>35%</b>	<b>37%</b>	<b>38%</b>	<b>31%</b>	<b>29%</b>	<b>25%</b>	<b>17%</b>	<b>17%</b>
	<i>63%</i>	<i>59%</i>	<i>53%</i>	<i>53%</i>	<i>50%</i>	<i>49%</i>	<i>33%</i>	<i>35%</i>
	34%	41%	42%	41%	34%	36%	24%	25%
Medical Science	<b>29%</b>	<b>34%</b>	<b>26%</b>	<b>25%</b>	<b>24%</b>	<b>24%</b>	<b>18%</b>	<b>18%</b>
	<i>62%</i>	<i>57%</i>	<i>55%</i>	<i>48%</i>	<i>53%</i>	<i>54%</i>	<i>41%</i>	<i>35%</i>
	13%	24%	25%	23%	32%	25%	24%	13%

### ***Types of impacts reflected by Wikipedia citations***

Assessing the strength of correlation between Wikipedia citations and academic citations can give insights into the extent to which they reflect the same types of impact. In particular, a strong positive correlation would suggest that they reflect similar types of impact, whereas weak correlations would

suggest that they either reflect different types of impact or that Wikipedia citations are essentially random and do not reflect any type of impact. Correlations were calculated separately for each year because citations take time to appear (e.g., two years or more in some fields) and so mixing data from different years can artificially inflate correlation coefficients (Fairclough & Thelwall, 2015).

There are significant, but low, Spearman correlations between Scopus citations and Wikipedia citations to articles in all fields and years (Table 4), suggesting that Wikipedia citations do not directly reflect academic research impact. The statistically significant positive correlations give evidence that Wikipedia citations are not purely random but do relate to academic impact in some way. This could occur if at least some of the Wikipedia citations are created for academic reasons, even if not all are. The weakness of the correlation coefficients is likely to be due to their relative sparseness (e.g., Table 1) and so this does not prove that Wikipedia citations do not reflect academic impact. The importance of the numbers of citations for the correlation coefficient is clear because the highest correlations between Scopus and Wikipedia citations were found for longer time periods 2005-2009 in all fields, with the additional citation data making the Spearman correlation tests more powerful. The lowest Spearman correlations mostly occurred for 2011-2012 (except for Biochemistry, Microbiology, Education and Philosophy), for the same reason.

**Table 4.** Spearman correlations between Wikipedia citations and Scopus citations to articles by year and field.

Scopus category	2005	2006	2007	2008	2009	2010	2011	2012
Astronomy and Astrophysics	.143**	.162**	<b>.170**</b>	.141**	.144**	.122**	<i>.098**</i>	.113**
Biochemistry and Genetics	<i>.168**</i>	.182**	.203**	.194**	<b>.326**</b>	.314**	.249**	.229**
Computer Science	<b>.139**</b>	.109**	.112**	.105**	.135**	.136**	<i>.071**</i>	.110**
Environmental Science	.127**	.140**	.102**	.127**	<b>.162**</b>	.130**	<i>.100**</i>	.135**
Internal Medicine	.170**	.161**	.175**	.164**	<b>.196**</b>	.166**	.173**	.133**
Microbiology (Medicine)	.159**	<b>.189**</b>	.117**	.133**	.114**	.172**	.147**	.144**
Pharmacology and Pharmaceutics	.165**	<b>.202**</b>	.169**	.176**	.172**	.160**	.114**	.134**
Surgery	.103**	.080**	.075**	.102**	<b>.116**</b>	.079**	<i>.052**</i>	.026
Business, Management	-.028	-.001	.026	.064**	<b>.130**</b>	.096**	.085**	<i>.054**</i>
Education	.090**	.096**	.049**	<b>.103**</b>	<i>.040*</i>	.089**	.054**	.079**
Psychology (clinical)	<b>.208**</b>	.201**	.192**	.168**	.155**	.095**	<i>.093**</i>	.114**
Political Science and Int. Relations	.174**	<b>.177**</b>	.106**	.137**	.062**	.096**	.076**	<i>.034*</i>
History	.134**	<b>.146**</b>	.144**	.090**	.098**	.089**	<i>.044*</i>	.065**
Literature	.121**	.042	.122**	<b>.127**</b>	.060**	.069**	<i>.035*</i>	-.010
Music, Visual Arts and Performing Arts	<b>.192**</b>	.151**	.114**	.102**	.078**	.072**	.061**	.025
Philosophy	.108**	.084**	<b>.151**</b>	.064**	<i>.062**</i>	.093**	.080**	.011
All fields	.078**	.085**	.084**	.087**	<b>.101**</b>	.100**	<i>.077**</i>	.086**

\* Significant at the p = 0.05 level. \*\* Significant at the p = 0.01 level. Bold: Highest correlation coefficients. Italic: Lowest correlation coefficients.

Google Books is the best source of academic-related citations with which to compare Wikipedia citations to monographs. The correlations are generally higher between Wikipedia citations and Google Books citations than between Wikipedia citations and Scopus citations (excluding in Business for 2005 and 2012). This suggests that the value of being cited in Wikipedia (e.g., the cultural or educational value) may be better reflected in book citations than in article citations, although the strength of the correlation may be partly due to there being more citations to books from Google Books than from Scopus. The correlations between Wikipedia citations and Google Books citations are, perhaps surprisingly, consistently highest in Physics and Astronomy, ranging from 0.736 in 2005 to 0.357 in 2012. This category included many introductory science books, such as "*The nature of space and time*" by Stephen Hawking and Roger Penrose (Wiki. cites: 21 and GB cites: 20) and "*Physical foundations of cosmology*" by V. Mukhanov (Wiki. cites: 35 and GB cites: 41), that tended to be cited in both Wikipedia articles and other science books. Another reason for stronger correlations in this field could be that Wikipedia seems to have particularly broad coverage of popular science topics, such as cosmology and astronomy, providing more Wikipedia citations for the correlation. The higher correlations for older books probably reflect citations increasing over the years.

There are stronger correlations between Wikipedia citations and academic syllabus mentions than between Wikipedia citations and Scopus citations in most fields, suggesting that Wikipedia citations are more closely related to educational benefits than to narrower scholarly benefits.

**Table 5.** Spearman correlations between Wikipedia citations with Google books, Scopus citations and academic syllabus mentions by year and broad field.

Scopus category	No. of Monographs	Correlations between Wikipedia citations and Google Books citations (bold)							
		<i>Correlations between Wikipedia citations and syllabus mentions (italic)</i>							
		Correlations between Wikipedia citations and Scopus citations							
		2005	2006	2007	2008	2009	2010	2011	2012
Arts and Humanities	3,762	<b>.354**</b>	<b>.371**</b>	<b>.271**</b>	<b>.310**</b>	<b>.326**</b>	<b>.327**</b>	<b>.271**</b>	<b>.337**</b>
		<i>.267**</i>	<i>.196**</i>	<i>.143**</i>	<i>.236**</i>	<i>.270**</i>	<i>.151**</i>	<i>.319**</i>	<i>.305**</i>
		.106*	.006	-.001	.105*	.100*	.134**	.200**	.135**
Social Sciences	3,807	<b>.394**</b>	<b>.337**</b>	<b>.278**</b>	<b>.254**</b>	<b>.303**</b>	<b>.301**</b>	<b>.354**</b>	<b>.321**</b>
		<i>.322**</i>	<i>.286**</i>	<i>.181**</i>	<i>.170**</i>	<i>.170**</i>	<i>.240**</i>	<i>.235**</i>	<i>.218**</i>
		.129**	.001	.070	.048	.081	.034	.073	.135**
Psychology	1,680	<b>.360**</b>	<b>.431**</b>	<b>.449**</b>	<b>.278**</b>	<b>.256**</b>	<b>.272**</b>	<b>.276**</b>	<b>.335**</b>
		<i>.278**</i>	<i>.250*</i>	<i>.197*</i>	<i>.198**</i>	<i>.390**</i>	<i>.155*</i>	<i>.310**</i>	<i>.211**</i>
		.123	.058	.095	.168*	.090	.015	.190**	.264**
Business and Management	1,099	<b>.203*</b>	<b>.270**</b>	<b>.229**</b>	<b>.262**</b>	<b>.208*</b>	<b>.250**</b>	<b>.232**</b>	<b>.171*</b>
		<i>.050</i>	<i>.129</i>	<i>.140</i>	<i>.329**</i>	<i>.303**</i>	<i>.186*</i>	<i>.147*</i>	<i>.010</i>
		.217*	0.046	.220**	.057	.173*	.156*	.213**	.219*
Biochemistry, Genetics and Molecular Bio.	1,351	<b>.599**</b>	<b>.388*</b>	<b>.409**</b>	<b>.315**</b>	<b>.261**</b>	<b>.282**</b>	<b>.236**</b>	<b>.323**</b>
		<i>.234</i>	<i>.251**</i>	<i>.176*</i>	<i>.250**</i>	<i>.245**</i>	<i>.205**</i>	<i>.252**</i>	<i>.216**</i>
		.142	.173*	.148	.223**	.015	.250**	.189**	.281**
Computer Science	1,572	<b>.359**</b>	<b>.407**</b>	<b>.421**</b>	<b>.305**</b>	<b>.256**</b>	<b>.253**</b>	<b>.322**</b>	<b>.179*</b>
		<i>.379**</i>	<i>.395**</i>	<i>.203**</i>	<i>.172*</i>	<i>.225**</i>	<i>.172**</i>	<i>.415**</i>	<i>.338**</i>
		.058	.056	.096	.033	.112	.120	.226**	.094
Environmental Science	1,472	<b>.162</b>	<b>.120</b>	<b>.295**</b>	<b>.260**</b>	<b>.241**</b>	<b>.125</b>	<b>.364**</b>	<b>.292**</b>
		<i>.079</i>	<i>.193*</i>	<i>.199*</i>	<i>.119</i>	<i>.048</i>	<i>.066</i>	<i>.149*</i>	<i>.053</i>
		.039	.031	.048	.056	.132	.063	.302**	.073

Physics and Astronomy	1,132	<b>.736**</b> .315* .231	<b>.619**</b> .116* .190	<b>.468**</b> .165* .087	<b>.586**</b> .292** .166	<b>.507**</b> .302** .134	<b>.474**</b> .331** .265**	<b>.370**</b> .246** .310**	<b>.357**</b> .254** .159*
Medicine+	2,860	<b>.343**</b> .231** .125	<b>.238**</b> .237** .111	<b>.326**</b> -.006 .059	<b>.344**</b> -.103 .009	<b>.244**</b> .202** .127**	<b>.226**</b> .079 .029	<b>.237**</b> .138** .108*	<b>.264**</b> .189** .082*
All Fields	18,735	<b>.406**</b> .287** .065**	<b>.369**</b> .248** .004	<b>.300**</b> .163** .025	<b>.351**</b> .226** .032	<b>.314**</b> .234** .081**	<b>.321**</b> .187** .091**	<b>.333**</b> .246** .208**	<b>.315**</b> .219** .161**

\* Significant at the  $p = 0.05$  level. \*\* Significant at the  $p = 0.01$  level. + Immunology and Microbiology and Pharmacology and Pharmaceutics books were merged with Medicine to have more data for the correlation analysis.

## Discussion

This study has a number of limitations. It does not cover all research areas and some areas may be particularly suited to, or alien to, Wikipedia. In addition, the choice of English language Scopus-indexed articles is an important restriction. Presumably, articles in other languages and those that are not in Scopus are less likely to be cited. The biases are probably substantially stronger for the monographs analysed because Scopus indexes relatively few academic books. Moreover, the automatic syllabus search method only captures a subset of the citations from academic course syllabi indexed by the Bing search engine and is likely to miss many results from academic syllabi that are not accessible in the open web.

An important limitation for interpretations of the results is that whilst Scopus citations reflect academic impact and syllabus mentions reflect educational impact, it is not clear what type of impact is reflected by citations from Google Books. Google does not reveal details about the extent of its coverage of different types of books and in any case only books that cite academic research are relevant here (e.g., few novels cite anything). Presumably, some of the citing books would be monographs and so their citations would reflect scholarly impact. Others would be textbooks, with their citations reflecting educational impact. In addition, there may be books aimed at a specific professional audience (e.g., *How to Improve Your Leadership and Management Skills*), reflecting commercial or societal impacts. These may overlap with educational impacts because books that are useful for a professional audience may also be added to relevant course reading lists. Some books citing academic research may also support the health and wellbeing (*Hardcore Self Help: F\*\*k Anxiety: Volume 1*) or life skills (e.g., *22 Things a Woman Must Know: If She Loves a Man With Asperger's Syndrome*) of the general population, reflecting societal impact. Some citing books will be popular works on the arts (e.g., *Great Film Directors A-Z*) and humanities (e.g., *A Short History of England*), reflecting art and cultural impacts. Popular books about science (e.g., *A Brief History of Time*) perhaps encompass both educational and cultural impacts. Because of these differences, Google Books citations probably tend reflect different types of impacts, depending on the cited disciplinary area.

Overall, Scopus-indexed academic monographs tend to be cited by Wikipedia considerably more often than are Scopus-indexed scientific articles (as confirmed by Table 6). Moreover, the correlations between Wikipedia citations and Google Books citations for monographs are much higher than the correlations between Wikipedia citations and Scopus citations for articles in all fields, suggesting that Wikipedia citations reflect a book type of impact rather than a type of article impact in all fields.



**Table 6.** A comparison between Wikipedia citations to monographs and to articles across broad fields.

Document Type	Monographs			Articles		
	Number	% with one or more Wiki. cites	Correlations: Wiki-GB Wiki-Scopus	Number	% with one or more Wiki. cites	Correlations: Wiki.-Scopus
Broad fields						
Arts and Humanities	3,762	48.0%	<b>.251**</b> .096**	69,106	7.2%	.117**
Social Sciences	6,586	34.5%	<b>.341**</b> 0.53**	77,237	4.7%	.105**
Sciences	5,527	27.0%	<b>.350**</b> .138**	79,430	4.5%	.179**
Medical Science	2,860	22.8%	<b>.280**</b> .068**	76,555	3.6%	.157**

An analysis of the monographs in each subject area with the most Wikipedia citations but no Scopus citations (Table 7) can give insights into the types of books that provide particular value to Wikipedia even though that value would not be recognised by traditional citation counts. There was only one case of a book with many apparently unmerited Wikipedia citations. This book had been added to the biography sections of hundreds of Serbo-Croatian Wikipedia pages related to Italy, presumably by a single enthusiastic editor. Three themes were common in the remaining books. Popular topics, such as astronomy, dinosaurs, Intelligence Quotient (IQ) tests, illegal drugs, and war, appear to have attracted a non-academic readership that included Wikipedia editors. Thus, their Wikipedia citations probably reflect wider popular uptake for these books. Textbooks and books that can be read by students also generated Wikipedia citations, perhaps from students learning about the topic and updating Wikipedia, or from their instructors ensuring that relevant pages have good citations. Reference works and reviews were also prominent. Even though academics tend to cite review articles more than other articles, this tendency may not extend as much to books, especially if they do not have a research focus. These last two types of citations may reflect educational impact rather than scholarly impact.

**Table 7.** Comments on the two books with no Scopus citations and the most Wikipedia citations in each subject area.

Scopus category	Monographs	Wiki cites	GB cites	Comment
Physics and Astronomy	Understanding variable stars	70	12	Review aimed at amateurs and undergraduates
Physics and Astronomy	Exploring ancient skies: An encyclopedic survey of archaeoastronomy	70		Review aimed at students and scholars – mainly relevant in archaeology
Pharmacology and Pharmaceutics	Pharmacology and abuse of cocaine, amphetamines, ecstasy and related designer drugs: A comprehensive review on their mode of action, treatment of abuse and intoxication	16	0	Review of interest to non-pharmacists amateurs because of illicit drug focus
Pharmacology and Pharmaceutics	The engines of Hippocrates: From the dawn of medicine to medical and	6	1	Reference guide

	pharmaceutical Informatics			
Immunology and Microbiology	Sulphate-reducing bacteria: Environmental and engineered systems	9	12	Review
Immunology and Microbiology	Control of innate and adaptive immune responses during infectious diseases	7	0	Essentially an encyclopaedia of infectious diseases
Medicine	International exposure: Perspectives on modern European pornography, 1800-2000	39	29	Popular topic (pornography) with a humanities orientation
Medicine	Principles of diabetes mellitus	37	0	Comprehensive textbook and reference
Environmental Science	Introduction to planetary science: The geological perspective	49	6	Comprehensive textbook
Environmental Science	Bernissart dinosaurs and early Cretaceous terrestrial ecosystems	48	0	Popular topic (dinosaurs)
Computer Science	Multiagent systems: Algorithmic, Game-Theoretic, and logical foundations	46	33	Textbook and reference
Computer Science	Algorithmic game theory	36	0	Comprehensive introduction, students part of audience
Biochemistry, Genetics and Molecular Biology	Handbook of behavior genetics	16	0	Textbook
Biochemistry, Genetics and Molecular Biology	Determining mycotoxins and mycotoxigenic fungi in food and feed	15	1	Comprehensive review
Business and Management	The innovation journey of Wi-Fi: The road to global success	7	6	Popular treatment of technical topic
Business and Management	Handbook Of trust research	5	0	Handbook - so a reference work
Psychology	What intelligence tests miss: The psychology of rational thought	32	30	Popular topic (IQ tests)
Psychology	What is intelligence?: Beyond the Flynn effect	23	29	Popular topic (IQ tests)
Social Sciences	Puritan village: The formation of a New England town	60	2	Popular topic (American settlers) and famous book, reviewed in Time
Social Sciences	The impact of European integration on political parties: Beyond the permissive consensus	51	6	Comprehensive coverage, students can read.
Arts and Humanities	Nature and history in modern Italy	603	8	Almost exclusively Serbo-Croatian Wikipedia biography citations
Arts and Humanities	Hell's islands: The untold story of Guadalcanal	135	5	Popular topic (World War 2)

There are considerable language differences in the proportions of Wikipedia citations to monographs between the main languages (Table 9 in the appendix). Over a half of the citations are from English language Wikipedia articles ([en.wikipedia.org/wiki/](http://en.wikipedia.org/wiki/) or [en.m.wikipedia.org](http://en.m.wikipedia.org)) and the remainder are mainly in German ([de.wikipedia.org/wiki/](http://de.wikipedia.org/wiki/)), French ([fr.wikipedia.org/wiki/](http://fr.wikipedia.org/wiki/)) and Spanish ([es.wikipedia.org/wiki/](http://es.wikipedia.org/wiki/)). Nevertheless, there are disciplinary differences in the proportions of the citing languages. English language Wikipedia citations were more numerous in Pharmacology and Pharmaceutics (about 73%) and Medicine (about 67%) than in Immunology and Microbiology and Biochemistry, Genetics (42%) and Molecular Biology (45%). Some English language monographs attracted Wikipedia citations in many different languages. For instance, "*Non-vascular interventional radiology of the abdomen*" by Ronald S. Arellano attracted 24 citations from non-English articles about multiple topics in 23 languages but only three citations from English language Wikipedia articles. Some citations were from also Wikipedia articles with the same topic but in different languages. One extreme example is the query *Barrett Stanberry "Vaccines for biodefense and emerging and neglected diseases" 2009 site:wikipedia.org/wiki/* that returned 31 matches, including Wikipedia entries on "*Dengue fever*" or "*Dengue*" in English, Ukrainian, Filipino, Persian, Japanese, Romanian, Greek, Italian and several other languages. Presumably one of these was the original article and the remainder were copied and translated whole or in part (Hale, 2015; Liao, 2014). Hence, some high Wikipedia citation counts may represent international copying and translation rather than a wide variety of different contributions.

More review articles (as labelled by Scopus) had at least one Wikipedia citation (7.7%) than did other articles (4.4%) across all fields (Figure 2). In Science and Medicine (except for astronomy and surgery) the proportion of review papers with one or more Wikipedia citation was at least double that of articles, indicating that reviews are more likely to be cited in Wikipedia articles. For instance, in Environmental Sciences 10% of review papers had at least one Wikipedia citation, whereas only 3% of other articles had one or more Wikipedia citation. In Social Science and the Arts and Humanities this difference is much lower, however. Review articles are usually longer than research articles and are perhaps more accessible to non-experts than are typical articles, and so may be naturally more useful for an encyclopaedia, perhaps for definitions, descriptions and background information (see: Bould et al., 2014).

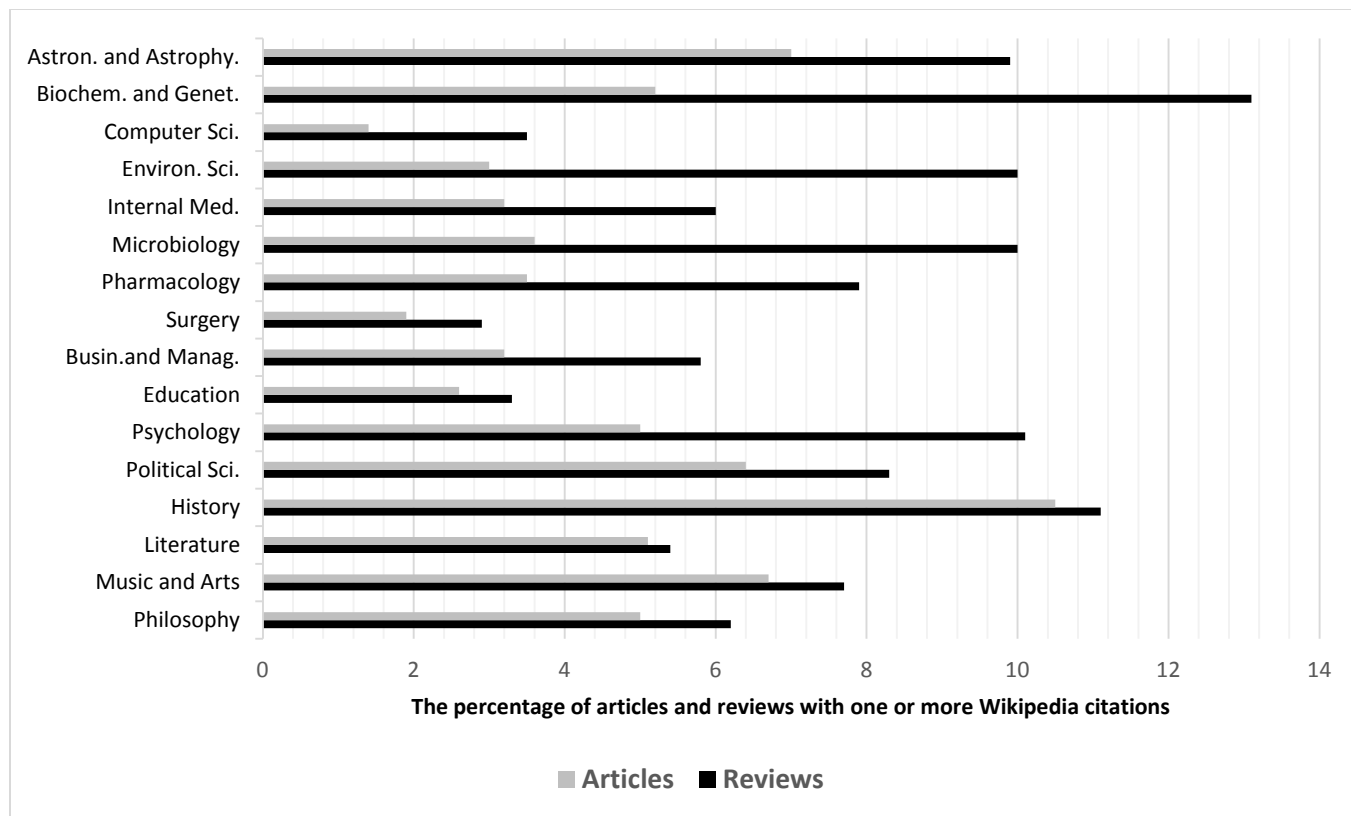


Figure 2. The Percentage of Scopus reviews or other articles with at least one Wikipedia citation.

## Conclusions

In answer to the first research question, only 5% of the Scopus articles had at least one Wikipedia citation, although 8% of review articles did, whereas 33% of the academic monographs had attracted one or more Wikipedia citations. This shows that Wikipedia citations to academic publications are much more common for books (see also Luyt & Tan, 2010; Halfaker & Taraborelli, 2015) rather than for articles (Priem, Piwowar, & Hemminger, 2012; Lin & Fenner, 2014). Moreover, only about 0.7% of the articles investigated had a Wikipedia citation but no Scopus citations whereas 22% of the monographs had a Wikipedia citation but no Scopus citations. Hence, indicators generated from Wikipedia citations are likely to be more useful for the impact assessment of books than for articles. For monographs, the figure of 33% being cited seems to be enough for Wikipedia citations to be useful for most research evaluation purposes, especially given that the numbers were not substantially smaller than those for the other sources checked. For articles, the percentage of Wikipedia-cited articles is too small to be worth calculating routinely for individual papers, with some exceptions. It would still be possible to compare large groups of papers based on the proportion attracting Wikipedia citations. For example, this might be used by a research group to support a claim that their work had a particularly strong impact in education. It would also be possible to use Wikipedia citations to identify individual articles that had made a strong contribution to education or popular knowledge even though they had not been cited much in the traditional literature.

In answer to the second research question, the weak but significant correlations between the Scopus citations and Wikipedia citations for most fields and years suggest that Wikipedia citations only loosely

reflect scientific impact. Higher positive correlations between Wikipedia citations and Google Books citations in almost all years and fields (Table 4), suggests that Wikipedia citations reflect more of a book type of impact (e.g., educational or cultural) than traditional scholarly impact. The moderate or strong correlations between Wikipedia citations and both Google Books citations and academic syllabus mentions in most fields are consistent with Wikipedia citations reflecting educational value, at least in part. Presumably, textbooks, introductory science books, histories, biographies, novels and poetry may be cited particularly often in encyclopaedia articles. It may be that Wikipedia citations also partially reflect other types of impact, such as value for popular science, but this was not tested. Hence, Wikipedia citations can be regarded as a particularly helpful quantitative indicator in most arts and humanities and in some areas of the social sciences, although the precise nature of the impact that they reflect probably varies by field.

In answer to the last research question, there were considerable disciplinary differences in the extent to which academic publications were cited in Wikipedia. Monographs were cited particularly often in the arts and humanities (48%) and in the social sciences (39%), probably due to the cultural or educational values of the books that were targeted at, or accessible to, students or a wider public. Variations between broad subject areas are also evident in the correlations between Wikipedia citations with Google books, Scopus citations and academic syllabus mentions and so in some areas Wikipedia citations may be closer to reflecting scholarly impact. For instance, the correlations between Wikipedia citations and the Google Books citations are high in Physics and Astronomy, perhaps due to many citations in popular Wikipedia entries about astronomy.

Overall, the results suggest that Wikipedia citations are a useful new source of evidence about the impacts of books, and probably their non-scholarly impacts, especially in the arts and humanities but also in the social sciences and elsewhere. This takes advantage of their open nature, in contrast to traditional academic citations, because any person with internet access can create and edit them. Wikipedia citations can also be used to compare groups of articles for non-scholarly impacts and to identify individual articles that make a substantial non-scholarly contribution. As with almost all alternative indicators (Wouters & Costas, 2012), however, the relative ease with which they can be manipulated means that they should not be used for research evaluations with stakeholders that would have an interest in the evaluations. They could still be used for self-evaluations and informal analyses as well as for more theoretical purposes, such as investigations into science dissemination.

## References

- Abramo, G., & D'Angelo, C.A. (2015). The VQR, Italy's second national research assessment: Methodological failures and ranking distortions. *Journal of the Association for Information, Science and Technology*. doi: 10.1002/asi.23323
- Adie, E., Roe, W. (2013). Altmetric: Enriching scholarly content with article-level discussion and metrics. *Learned Publishing*, 26: 11–17. [http://figshare.com/articles/Enriching\\_scholarly\\_content\\_with\\_article\\_level\\_discussion\\_and\\_metrics/105851](http://figshare.com/articles/Enriching_scholarly_content_with_article_level_discussion_and_metrics/105851).
- Aibar, E., Lladós-Masllorens, J., Meseguer-Artola, A., Minguillón, J., & Lerga, M. (2015). Wikipedia at university: What faculty think and do about it. *Electronic Library*, 33(4), 668-683.
- Ball, R. & Tunger, D. (2006). Science indicators revisited – Science Citation Index versus SCOPUS: A bibliometric comparison of both citation databases. *Information Services & Use*, 26 (4), 293-301.
- Bar-Ilan, J. (2004). The use of Web search engines in information science research. *Annual review of information science and technology*, 38(1), 231-288.

- Bar-Ilan, J. (2008). Which h-index? - A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257-271.
- Bayliss, G. (2013). Exploring the cautionary attitude toward Wikipedia in higher education: Implications for higher education institutions. *New Review of Academic Librarianship*, 19(1), 36-57.
- Bould, M. D., Hladkowitz, E. S., Pigford, A. E., Ufholz, L., Postonogova, T., Shin, E., & Boet, S. (2014). References that anyone can edit: Review of Wikipedia citations in peer reviewed health science literature. *BMJ (Online)*, 348, g1585. doi: <http://dx.doi.org/10.1136/bmj.g1585>
- Brazzeal, B. (2011). Citations to Wikipedia in chemistry journals: A preliminary study. *Issues in Science and Technology Librarianship*, 67. doi: 10.5062/F4057CV7
- Chen, H. -L. (2010). The perspectives of higher education faculty on Wikipedia. *Electronic Library*, 28(3), 361-373.
- Chen, H.-L. (2009). The use and sharing of information from Wikipedia by high-tech professionals for work purposes. *Electronic library*, 27(6), 893–905
- Chesney, T. (2006). An empirical examination of Wikipedia's credibility. *First Monday*, 11(11). <http://journals.uic.edu/ojs/index.php/fm/article/view/1413/1331>
- Cohen, N. (2007). A history department bans citing Wikipedia as a research source, *New York Times* (21 February), <http://www.nytimes.com/2007/02/21/education/21wikipedia.html>
- de Winter, J. C. F., Zadpoor, A. A., & Dodou, D. (2014). The expansion of Google Scholar versus Web of Science: A longitudinal study. *Scientometrics*, 98(2), 1547-1565.
- Denning, P., Horning, J., Parnas, D., & Weinstein, L. (2005). Wikipedia risks. *Communications of the ACM*, 48(12), 152.
- Devgan, L., Powe, N., Blakey, B., & Makary, M. (2007). Wiki-Surgery? Internal validity of Wikipedia as a medical and surgical reference. *Journal of the American College of Surgeons*, 205(3, supplement), S76–S77. doi:10.1016/j.jamcollsurg.2007.06.190
- Dooley, P. L. (2010). Wikipedia and the two-faced professoriate. Paper presented at the Proceedings of WikiSym 2010 - the 6th International Symposium on Wikis and Open Collaboration, Gdansk; Poland; 7 July-9 July. doi:10.1145/1832772.1832803
- Eijkman, H. (2010). Academics and Wikipedia: Reframing web 2.0+as a disruptor of traditional academic power-knowledge arrangements. *Campus-Wide Information Systems*, 27(3), 173-185.
- Elsevier (2013a). International Comparative Performance of the UK Research Base - 2013. <https://www.gov.uk/government/publications/performance-of-the-uk-research-baseinternational-comparison-2013>
- Elsevier (2013b). Elsevier announces its Scopus book titles expansion program. <http://www.elsevier.com/about/press-releases/science-and-technology/elsevier-announces-its-scopus-book-titles-expansion-program>
- Fairclough, R. & Thelwall, M. (2015). The influence of time and discipline on the magnitude of correlations between citation counts and quality scores. *Journal of Informetrics*, 9(3), 529–541.
- Fallis, D. (2008). Toward an epistemology of Wikipedia. *Journal of the American Society for Information Science and Technology*, 59(10), 1662-1674.
- Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438(7070), 900–901. doi:10.1038/438900a
- Gorman, G. E. (2007). A tale of information ethics and encyclopædias; or, is Wikipedia just another internet scam? *Online Information Review*, 31(3), 273-276.
- Haigh, C. A. (2011). Wikipedia as an evidence source for nursing and healthcare students. *Nurse Education Today*, 31(2), 135-139.
- Hale, S.A. (2015) Cross-language Wikipedia editing of Okinawa, Japan. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '15), New York: ACM Press (pp. 1-8).

- Halfaker, A., & Taraborelli, D. (2015). Wikipedia scholarly article citations. <http://dx.doi.org/10.6084/m9.figshare.1299540>
- Harzing, A. (2014). A longitudinal study of Google Scholar coverage between 2012 and 2013. *Scientometrics*, 98(1), 565-575.
- Head, A. J., & Eisenberg, M. B. (2010). How today's college students use Wikipedia for course-related research. *First Monday*, 15(3).
- Heilman, J. M., & West, A. G. (2015). Wikipedia and medicine: Quantifying readership, editors, and the significance of natural language. *Journal of Medical Internet Research*, 17(3) doi:10.2196/jmir.4069
- Heilman, J.M., Kemmann, E., Bonert, M., Chatterjee, A., Ragar, B., Beards, G.M., . . . Laurent, M.R. (2011). Wikipedia: A key tool for global public health promotion. *Journal of Medical Internet Research*, 13(1), e14. doi:10.2196/jmir.1589
- Hughes, B., Joshi, I., Lemonde, H., & Wareham, J. (2009). Junior physician's use of web 2.0 for information seeking and medical education: A qualitative study. *International Journal of Medical Informatics*, 78(10), 645-655.
- Jacsó, P. (2011). Google Scholar duped and deduped - the aura of "robometrics". *Online Information Review*, 35(1), 154-160.
- Judd, T., & Kennedy, G. (2011). Expediency-based practice? Medical students' reliance on Google and Wikipedia for biomedical inquiries. *British Journal of Educational Technology*, 42(2), 351-360.
- Jullien, N. (2012). What we know about Wikipedia. A review of the literature analyzing the project(s). [https://papers.ssrn.com/sol3/Delivery.cfm/SSRN\\_ID2053597\\_code728676.pdf?abstractid=2053597&mirid=2](https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID2053597_code728676.pdf?abstractid=2053597&mirid=2)
- Khabsa, M., & Giles, C. L. (2014). The number of scholarly documents on the public web. *PLoS ONE*, 9(5): e93949. doi:10.1371/journal.pone.0093949.
- Kim, J. Y., Gudewicz, T. M., Dighe, A. S., & Gilbertson, J. R. (2010). The pathology informatics curriculum wiki: Harnessing the power of user-generated content. *Journal of Pathology Informatics*, 1. doi:10.4103/2153-3539.65428
- Knight, C., & Pryke, S. (2012). Wikipedia and the university, a case study. *Teaching in Higher Education*, 17(6), 649-659.
- Kousha, K. & Thelwall, M. (2015). An automatic method for assessing the teaching impact of books from online academic syllabi. *Journal of the Association for Information Science and Technology*. doi: 10.1002/asi.23542
- Kousha, K. & Thelwall, M. (in press 2015). Patent citation analysis with Google. *Journal of the Association for Information Science and Technology*. <http://www.scit.wlv.ac.uk/~cm1993/papers/PatentCitationAnalysisWithGooglePreprint.pdf>
- Kousha, K., & Thelwall, M. (2007a). Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology*, 58(7), 1055-1065.
- Kousha, K., & Thelwall, M. (2008). Assessing the impact of disciplinary research on teaching: An automatic analysis of online syllabuses. *Journal of the American Society for Information Science and Technology*, 59(13), 2060-2069.
- Kousha, K., & Thelwall, M. (2009). Google Book Search: Citation analysis for social science and the humanities. *Journal of the American Society for Information Science and Technology*, 60(8), 1537-1549.
- Kousha, K., & Thelwall, M. (2014). An automatic method for extracting citations from Google Books. *Journal of the Association for Information Science and Technology*. 66(2), 309-320.

- Kousha, K., Thelwall, M., & Rezaie, S. (2011). Assessing the citation impact of books: The role of Google Books, Google Scholar, and Scopus. *Journal of the American Society for Information Science and Technology*, 62(11), 2147-2164.
- Kovic, I., Lulic, I. & Brumini, G. (2008). Examining the medical blogosphere: An online survey of medical bloggers. *Journal of Medical Internet Research*, 10(3), e28.
- Kryl, D., Allen, L., Dolby, K., Sherbon, B., & Viney, I. (2012). Tracking the impact of research on policy and practice: investigating the feasibility of using citations in clinical guidelines for research evaluation. *BMJ Open*, 2(2), e000897. doi: 10.1136/bmjopen-2012-000897
- Labbé, C., & Labbé, D. (2013). Duplicate and fake publications in the scientific literature: How many SCIdgen papers in computer science? *Scientometrics*, 94(1), 379-396.
- Laurent, M. R., & Vickers, T. J. (2009). Seeking health information online: Does Wikipedia matter? *Journal of the American Medical Informatics Association*, 16(4), 471-479.
- Leydesdorff, L., & Felt, U. (2012). Edited volumes, monographs and book chapters in the Book Citation Index (BKCI) and Science Citation Index (SCI, SoSCI, A&HCI). *Journal of Scientometric Research*, 1(1), 28-34.
- Liao, H. (2014). *The cultural politics of user-generated encyclopedias: Comparing Chinese Wikipedia and Baidu*. Oxford, UK: University of Oxford.
- Lim, S. (2009). How and why do college students use Wikipedia? *Journal of the American Society for Information Science and Technology*, 60(11), 2189-2202.
- Lin, J. & Fenner, M. (2014). An analysis of Wikipedia references across PLOS publications. Paper presented at the altmetrics14: Expanding impacts and metrics An ACM Web Science Conference 2014 Workshop, 23- 26 June, Indiana University, Indiana, USA. <http://dx.doi.org/10.6084/m9.figshare.1048991>
- López-Cózar, E. D., Robinson-García, N., & Torres-Salinas, D. (2014). The Google Scholar experiment: How to index false papers and manipulate bibliometric indicators. *Journal of the Association for Information Science and Technology*, 65(3), 446-454.
- Luyt, B., & Tan, D. (2010). Improving Wikipedia's credibility: References and citations in a sample of history articles. *Journal of the American Society for Information Science and Technology*, 61(4), 715-722.
- Luzón, M. J. (2009). Scholarly hyperwriting: The function of links in academic weblogs. *Journal of the American Society for Information Science and Technology*, 60(1), 75-89.
- MacRoberts, M.H., & MacRoberts, B.R. (1996). Problems of citation analysis. *Scientometrics*, 36(3), 435-444.
- MacRoberts, M.H., & MacRoberts, B.R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science*, 40(5), 342-349.
- Meester, W. (2013). Towards a comprehensive citation index for the Arts & Humanities, *Research Trends*, 32. <http://www.researchtrends.com/issue-32-march-2013/towards-a-comprehensive-citation-index-for-the-arts-humanities/>
- Mehegan, D. (2006). Bias, sabotage haunt Wikipedia's free world. *The Boston Globe*, February 12, 2006. [http://www.boston.com/news/nation/articles/2006/02/12/bias\\_sabotage\\_haunt\\_wikipedias\\_free\\_world/](http://www.boston.com/news/nation/articles/2006/02/12/bias_sabotage_haunt_wikipedias_free_world/)
- Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105-2125.
- Mesgari, M., Okoli, C., Mehdi, M., Nielsen, F. Å., & Lanamäki, A. (2015). "The sum of all human knowledge": A systematic review of scholarly research on the content of Wikipedia. *Journal of the Association for Information Science and Technology*, 66(2), 219-245. doi:10.1002/asi.23172
- Moed, H. F. (2005). *Citation analysis in research evaluation*. New York, New York: Springer.



- Narin, F., & Olivastro, D. (1992). Status report: Linkage between technology and science. *Research Policy*, 21(3), 237-249.
- Nielsen, F. A. (2007). Scientific citations in Wikipedia. *First Monday*, 12(8).  
<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1997/1872>
- Okoli, C., Mehdi, M., Mesgari, M., Nielsen, F. Å., & Lanamäki, A. (2014). Wikipedia in the eyes of its beholders: A systematic review of scholarly research on Wikipedia readers and readership. *Journal of the Association for Information Science and Technology*, 65(12), 2381-2403.
- Park, T. K. (2011). The visibility of Wikipedia in scholarly publications. *First Monday*, 16(8).  
<http://www.ojphi.org/ojs/index.php/fm/article/view/3492>
- Priem, J., Piwowar, H., & Hemminger, B. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. Retrieved from <http://arXiv.org/html/1203.4745v1>
- Rainie, L., & Tancer, B. (2007). Data memo. [http://www.pewinternet.org/files/old-media//Files/Reports/2007/PIP\\_Wikipedia07.pdf.pdf](http://www.pewinternet.org/files/old-media//Files/Reports/2007/PIP_Wikipedia07.pdf.pdf)
- Schmoch, U. (1993). Tracing the knowledge transfer from science to technology as reflected in patent indicators. *Scientometrics*, 26(1), 193-211.
- Scopus content coverage guide (2014).  
[https://www.elsevier.com/\\_\\_\\_data/assets/pdf\\_file/0007/69451/sc\\_content-coverage-guide\\_july-2014.pdf](https://www.elsevier.com/___data/assets/pdf_file/0007/69451/sc_content-coverage-guide_july-2014.pdf)
- Shema, H., Bar-Ilan, J., & Thelwall, M. (2014). Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics. *Journal of the Association for Information Science and Technology*, 65(5), 1018-1027.
- Soules, A. (2015). Faculty perception of Wikipedia in the California state university system. *New Library World*, 116(3-4), 213-226.
- Spoerri, A. (2007). What is popular on Wikipedia and why? *First Monday*, 12(4). <http://ojs-prod-lib.cc.uic.edu/ojs/index.php/fm/article/view/1765/1645>
- Stankus, T., & Spiegel, S. E. (2010). Wikipedia, Scholarpedia, and references to books in the brain and behavioral sciences: A comparison of cited sources and recommended readings in matching free online encyclopedia entries. *Science and Technology Libraries*, 29(1-2), 144-164.
- Sud, P. & Thelwall, M. (2014). Evaluating altmetrics. *Scientometrics*, 98(2), 1131-1143.
- Teplitskiy, M., Lu, G., & Duede, E. (2015). Amplifying the impact of Open Access: Wikipedia and the diffusion of science. <http://arxiv.org/abs/1506.07608>
- The Book Citation Index in Web of Science (2015).  
[http://wokinfo.com/products\\_tools/multidisciplinary/bookcitationindex/](http://wokinfo.com/products_tools/multidisciplinary/bookcitationindex/)
- Thelwall, M., & Kousha, K. (2008). Online presentations as a source of scientific impact?: An analysis of PowerPoint files citing academic journals. *Journal of the American Society for Information Science & Technology*, 59(5), 805-815.
- Thelwall, M. & Kousha, K. (2015). Web indicators for research evaluation. Part 1: Citations and links to academic articles from the Web. *El profesional de la información*, 24(5), 587-606. DOI: <http://dx.doi.org/10.3145/epi.2015.sep.08>
- Thelwall, M., & Maflahi, N. (2015). Guideline references and academic citations as evidence of the clinical value of health research. *Journal of the Association for Information Science and Technology*. doi: 10.1002/asi.23432
- Thomas, G. R., Eng, L., de Wolff, J. F., & Grover, S. C. (2013). An evaluation of Wikipedia as a resource for patient education in nephrology. *Seminars in Dialysis*, 26(2), 159-163. doi:10.1111/sdi.12059
- Thomson Reuters (2015). Book Citation Index.  
[http://wokinfo.com/products\\_tools/multidisciplinary/bookcitationindex/](http://wokinfo.com/products_tools/multidisciplinary/bookcitationindex/)

- Vaughan, L., & Shaw, D. (2003). Bibliographic and web citations: What is the difference? *Journal of the American Society for Information Science and Technology*, 54(14), 1313-1322.
- Vaughan, L., & Shaw, D. (2005). Web citation data for impact assessment: A comparison of four science disciplines. *Journal of the American Society for Information Science and Technology*, 56(10), 1075-1087.
- Web of Science core collection (2015).  
[http://wokinfo.com/products\\_tools/multidisciplinary/webofscience/](http://wokinfo.com/products_tools/multidisciplinary/webofscience/)
- Wouters, P., & Costas, R. (2012). Users, narcissism and control: Tracking the impact of scholarly publications in the 21st century. In: E. Archambault, Y. Gingras, V. Larivière (Eds) *Proceedings of the 17th International Conference on Science and Technology Indicators*. Montreal: Science-Metrix and OST (Vol. 2, pp. 487-497).
- Wouters, P., Thelwall, M., Kousha, K., Waltman, L., de Rijcke, S., Rushforth, A., & Franssen, T. (2015). The metric tide: Literature review (Supplementary report I to the independent review of the role of metrics in research assessment and management). The Higher Education Funding Council for England. doi: 10.13140/RG.2.1.5066.3520.  
[http://blogs.lse.ac.uk/impactofsocialsciences/files/2015/07/2015\\_metrictideS1.pdf](http://blogs.lse.ac.uk/impactofsocialsciences/files/2015/07/2015_metrictideS1.pdf)
- Yasseri, T., Spoerri, A., Graham, M., & Kertész, J. (2014). The most controversial topics in Wikipedia: A multilingual and geographical analysis. In: Fichman P., Hara N., editors, *Global Wikipedia: International and cross-cultural issues in online collaboration*. Lanham, MD: Rowman & Littlefield Education (pp. 21-44).

## Appendix

**Table 8.** The 20 articles that were cited most often in Wikipedia but that had no Scopus citations.

Article	General subject	Wiki citations	Scopus citations
Pierce, D. (2007). Forgotten faces: Why some of our cinema heritage is part of the public domain. <i>Film History: An International Journal</i> .	Film and media studies	82	0
Leonard, D. (2006). 'South Park' creators haven't lost their edge. <i>Fortune</i> .	Business	25	0
Van Dyke, C. (2010). The truth, the whole truth, and nothing but the truth: Robert Grosseteste on universals. <i>Journal of the History of Philosophy</i> .	Philosophy / History	22	0
Kizilov, M. (2007). Slaves, money lenders, and prisoner guards: The Jews and the trade in slaves and captives in the Crimean Khanate. <i>Journal of Jewish Studies</i> .	History / Cultural studies	19	0
Lachlan, R.B. (2009). Two new species of <i>Gnathothlibus</i> Wallengren from Fiji and Samoa and a new species of <i>Theretra</i> Hübner from New Guinea (Lepidoptera: Sphingidae). <i>Records of the Australian Museum</i> .	History / Museology	16	0
Tompkins, J. (2009). What's the deal with soundtrack albums? Metal music and the customized aesthetics of contemporary horror. <i>Cinema Journal</i> .	Film and media studies	16	0
Loktin, A.V. (2006). Kinematics of stars in the Pleiades open cluster. <i>Astronomy Reports</i> .	Astronomy	14	0
Potter, R. (2009). Obscene modernism and the trade in salacious books. <i>Modernism - Modernity</i> .	Music	14	0
Lewis, D. (2009). Matisse and Byzantium, or, mechanization takes command. <i>Modernism - Modernity</i> .	Music	13	0
Carver, A.F. (2005). Bruckner and the Phrygian mode. <i>Music and Letters</i> .	Music	12	0
Stoltzfus, B. (2005). Sartre, Nada, and Hemingway's African stories. <i>Comparative Literature Studies</i> .	Literature	11	0
Arrowsmith, R.R. (2011). The transcultural roots of modernism: Imagist poetry, Japanese visual culture, and the western museum system. <i>Modernism - Modernity</i> .	History / Cultural studies	10	0
Brands, H.W. (2012). Grant takes on the Klan. <i>American History</i> .	History	10	0
Steinmann, A. (2009). When did Herod the Great reign? <i>Novum Testamentum</i> .	History	10	0
Mockford, J. (2005). Before Lewis and Clark, Lt. Broughton's river of names: The Columbia river exploration of 1792. <i>Oregon Historical Quarterly</i> .	History	10	0
Carroll, B.D. (2012). Savages in the service of empire: Native American soldiers in Gorham's rangers, 1744-1762. <i>New England Quarterly-A Historical Review of New England Life and Letters</i> .	History	10	0
McIntyre, W.D. (2008). The expansion of the commonwealth and the criteria for membership. <i>The Round Table: The Commonwealth Journal of International Affairs</i> .	Political science	10	0
Mortimer, I. (2007). Henry IV's date of birth and the royal Maundy. <i>Historical Research</i> .	History	9	0
Sayers, J. (2009). A once 'Proud Prelate': An unidentified episcopal monument in Ely cathedral. <i>Journal of the British Archeological Association</i> .	Archaeology	9	0
Gusejnov, G. (2009). The linguistic aporias of Alexei Losev's mystical personalism. <i>Studies in East European Thought</i> .	Philosophy	9	0

**Table 9.** The language of the Wikipedia articles citing the sampled monographs by field.

Scopus category	Wikipedia languages*								
	en	de	fr	es	ru	pt	zh	it	Other
Arts and Humanities	51.1%	4.5%	4.7%	3.3%	2.6%	2.0%	1.7%	1.6%	28.5%
Social Sciences	60.4%	5.8%	3.2%	2.4%	1.7%	1.5%	2.0%	1.4%	21.6%
Psychology	63.1%	4.7%	2%	3.9%	2.8%	1.1%	1.4%	1.6%	19.4%
Business and Management	61.2%	7.6%	3.9%	2.4%	2.9%	1.4%	1.4%	0.8%	18.4%
Biochemistry, Genetics and Molecular Biology	44.9%	7.1%	2.1%	2.4%	3.5%	1.7%	2.5%	4.4%	31.4%
Computer Science	54.9%	7.0%	3.8%	3%	4.4%	3.1%	1.8%	2.4%	19.6%
Environmental Science	52.7%	6.8%	4.3%	2.8%	3.2%	1.4%	1.6%	1.7%	25.5%
Physics and Astronomy	58.6%	7.1%	3.3%	4%	2.9%	0.02%	2.6%	2.8%	18.7%
Medicine	66.7%	3.9%	2.2%	1.4%	1.6%	0.9%	1.1%	3.2%	19.0%
Immunology and Microbiology	42.5%	5.0%	3.4%	4.0%	1.1%	2.8%	0.6%	4.5%	36.1%
Pharmacology and Pharmaceutics	72.9%	1.2%	0.6%	0.7%	1.2%	0.2%	0.5%	15.3%	7.4%

\* <http://en.wikipedia.org/wiki> or <http://de.wikipedia.org/wiki> etc.