"The sum of all human knowledge":
A systematic review of scholarly research on
the content of Wikipedia

Mostafa Mesgari
John Molson School of Business, Concordia University, Montreal, Canada
mmesgari@jmsb.concordia.ca

Chitu Okoli
John Molson School of Business, Concordia University, Montreal, Canada
Chitu.Okoli@concordia.ca

Mohamad Mehdi
Computer Science, Concordia University, Montreal, Canada
mo_mehdi@encs.concordia.ca

Finn Årup Nielsen
DTU Compute, Technical University of Denmark, Kongens Lyngby, Denmark
faan@dtu.dk

Arto Lanamäki
Department of Information Processing Science, University of Oulu, Oulu, Finland
arto.lanamaki@oulu.fi

Abstract

Wikipedia might possibly be the best-developed attempt thus far of the enduring quest to gather all human knowledge in one place. Its accomplishments in this regard have made it an irresistible point of inquiry for researchers from various fields of knowledge. A decade of research has thrown light on many aspects of the Wikipedia community, its processes, and content. However, due to the variety of the fields inquiring about Wikipedia and the limited synthesis of the extensive research, there is little consensus on many aspects of Wikipedia's content as an encyclopedic collection of human knowledge. This study addresses the issue by systematically reviewing 110 peer-reviewed publications on Wikipedia content, summarizing the current findings, and highlighting the major research trends. Two major streams of research are identified: the quality of Wikipedia content (including comprehensiveness, currency, readability and reliability) and the size of Wikipedia. Moreover, we present the key research trends in terms of the domains of inquiry, research design, data source, and data gathering methods. This review synthesizes scholarly understanding of Wikipedia content and paves the way for future studies.

1

Keywords: Wikipedia, systematic literature review, encyclopedias, quality, comprehensiveness, currency, readability and style, reliability, accuracy, size of Wikipedia, featured articles, open content

Introduction

> Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge. That's what we're doing.
>
> — Jimmy Wales, founder of Wikipedia (Slashdot, 2004)

Wikipedia is one of the most striking emblems of the Web 2.0 era. Its broad encyclopedic content boldly attempts to encompass "all human knowledge", which is presently embodied in over 26 million articles across over 250 languages. Its enormous breadth with easy accessibility on the Web has made Wikipedia an essential source of information for many people with various information needs and purposes. Wikipedia is the sixth most-visited website globally with over 500 million readers each month[1]. Its rapid growth since 2001 and the widespread use of its content have made Wikipedia a focal point of inquiry for scholars from a wide variety of fields.

Despite the large volume of research on Wikipedia, and despite the existence of a number of literature reviews on Wikipedia research (for a review of reviews, see Okoli, Mehdi, Mesgari, Nielsen, & Lanamäki, 2012, sec. Literature reviews), there has as of yet been no comprehensive review that focuses specifically on the most quintessential aspect of the Wikipedia phenomenon: its encyclopedic content. Thus, our focus in this systematic literature review is not on the Wikipedia community, whether the contributors or readers, but on the product of Wikipedia—the encyclopedia articles themselves.

There are two major aspects of Wikipedia content that have not been adequately treated in past reviews (Jullien, 2012; Martin, 2010). First, perhaps the most researched question about Wikipedia since its foundation has been, "how good is it?" That is, is this encyclopedia built mostly by anonymous non-expert contributors a high-quality product? A major shortage in existing reviews has been the failure to distinguish the various aspects of "quality" that are appropriate for considering Wikipedia. In this review we distinguish studies on several important quality dimensions, including comprehensiveness, currency, readability and reliability (accuracy); we thus give a much more thorough perspective of Wikipedia's quality than has been previously conducted. A second major aspect of the content is the size of Wikipedia. No previous review has synthesized the studies on this important aspect of Wikipedia as we do here. In addition, we review studies on several other aspects of Wikipedia's content.

The contribution of this study is twofold. Firstly, it systematically reviews the extant research around the various aspects of Wikipedia content, identifies the main streams of research, and summarizes and represents the current state of knowledge along these streams. This provides the research community with informed insights about the extensively varied domain of inquiries on Wikipedia content, which can be used as a knowledge map for further inquiries; it also highlights the under-investigated areas of research that would be helpful in formulating new lines of study. Secondly, it analyzes the trends in all aspects of Wikipedia content research across streams of research, ranging from chronological trends to various trends in theory type, research design, and data-gathering methods employed by the research community. This may inform the design of studies to better capture unique aspects of Wikipedia content, and even point to underused methods that may encourage new designs and approaches to studying the Wikipedia phenomenon.

This present study reports on part of an extensive research project systematically reviewing the scholarly research on Wikipedia. This multi-year systematic literature review project, called "WikiLit"

---

[1] http://www.alexa.com/siteinfo/wikipedia.org; http://reportcard.wmflabs.org/

(http://wikilit.referata.com), extensively reviews scholarly research specifically examining Wikipedia, or using Wikipedia in research inquiry as a case or data source. The review project followed Okoli and Schabram's (2010) systematic review methodology, started with preparing a review protocol (Okoli & Schabram, 2009a), and continued with a systematic search process, practical screening, literature search, and data synthesis. For systematic search, we searched through 484 English-language scholarly databases available at Concordia University, Montreal (as of 2009) for studies with "Wikipedia", "Wikipedian", or "Wikipedians" in their title, abstract or keywords. During the practical screen, we examined each of the 2,678 unique items found to retain the peer-reviewed studies that examined Wikipedia as a significant subject. Several other search techniques were conducted to identify further articles. This process systematically identified and examined over 500 peer-reviewed scholarly studies including journal papers and doctoral theses published up until June 2011, as well as over 100 of the most influential conference papers. The details of the inclusion criteria and the review procedures are reported in an overview paper (Okoli et al., 2012, sec. Methodology). The identified studies have been coded, categorized, analyzed, and presented in an interactive website enabling visitors to observe and analyze the findings, and use it for their own research projects (WikiLit: http://wikilit.referata.com). We hope that the website also makes a meaningful and distinct contribution to the research community.

Through the WikiLit project, six main areas of inquiry about Wikipedia are identified: general Wikipedia studies, infrastructure, content, participation, readership, and corpus. This present study summarizes the findings and presents the trends concerning the content of Wikipedia; the other aspects of Wikipedia research are presented in other publications (Okoli, Mehdi, Mesgari, Nielsen, & Lanamäki, 2014; Okoli et al., 2012). The content category includes studies related to Wikipedia's encyclopedic content, its growth, its depth, breadth, and reliability, mainly focusing on the encyclopedia articles and the structure in which they are presented. Coding the topics of studies and continuously iterating between emerging topics and existing research, we identified the two main subtopics of Wikipedia content to be quality and size of Wikipedia, though some other various studies are represented as a third category. The content quality stream comes down to six major sub-streams examining antecedents of content quality, comprehensiveness, currency, readability, and reliability aspects of content quality, as well as featured articles (Wikipedia articles identified by the community as high-quality). Table 1 depicts these topic categories and the number of articles examining each one. The topic categories are not mutually exclusive, because a number of papers have looked into multiple aspects of Wikipedia content; thus the number of articles in sub-categories does not add up to that of the higher category. Research findings along each topic category follow.

Table 1. Categorization of Topics of Wikipedia Content Research

```
► Content (98)
   ► Quality (82)
      - Antecedents of quality (18)
      - Comprehensiveness (26)
      - Currency (7)
      - Featured articles (21)
      - Readability and style (13)
      - Reliability (30)
   - Size of Wikipedia (13)
   - Other content topics (9)
```

Number of studies in each topic are in parentheses

The paper is organized as follows: the next section presents the main research streams, and provides the findings in each of the streams, as well as in each of their sub-streams. Next, the trends in Wikipedia content research are briefly analyzed and related tables are provided in an appendix. Finally, the paper concludes with a summary of current understanding of various aspects of Wikipedia content research.

In considering the descriptions of the research studies, we must note that almost all the studies treat the English Wikipedia (see Table 16), and so the scope of findings should be considered accordingly. We explicitly note when studies treated a language version other than English.

Quality of Content

Quality of Wikipedia articles is one of the main concerns of the academic and user communities about Wikipedia, mainly due to the non-expert and openly participatory nature of Wikipedia development. Lewandowski and Spree (2011) found that Wikipedia results shown on search engines are quite dependent on the quality of articles; thus quality of articles has important direct consequences for its readership and existence on the Web. Hence, many researchers have investigated this important aspect of Wikipedia content. Such studies typically select a sample of Wikipedia articles and sometimes "manually" read and judge the quality, sometimes measure it using other objective criteria, and sometimes in comparison with other encyclopedias or other resources.

In this study, we differentiate between the concepts of objective expert evaluations of quality from readers' or contributors' subjective perceptions of quality. Whereas readers and Wikipedia contributors do have their own individual subjective perceptions of the credibility or quality of articles, we do not consider such perceptions sufficiently objective to be valid assessments of the actual or real quality of Wikipedia content. Thus, studies probing into such subjective quality perceptions are identified as Reader Perceptions of Credibility (Okoli et al., 2014) and Contributor Perceptions of Credibility (Okoli et al., 2012, sec. Contributor Perceptions of Credibility), which we present in other related publications. In this present review, we describe scholarly studies that explicitly sought out qualified subject experts to try to evaluate the quality of Wikipedia using objective standards.

Actually, "quality" is a multi-dimensional concept and studies have investigated its various aspects. More precisely, studies of Wikipedia's "quality" have looked into one or more of the following aspects: Reliability or accuracy (that is, absence of factual errors); comprehensiveness or breadth of coverage of subject matter, whether within an individual article or across multiple articles; currency or up-to-dateness of the article contents; and readability and quality of writing style. In addition to these precise quality topics, some articles have studied antecedents to these various aspects of quality. A final important group of articles investigated various aspects of featured articles, those articles vetted by the Wikipedia community as being of high quality.

Comprehensiveness

As envisioned by its founder, Wikipedia is purportedly aimed at incorporating all human knowledge within an encyclopedia (Slashdot, 2004), so comprehensiveness is always a major point of inquiry about Wikipedia and an important aspect of its quality in considering how much of the knowledge from different fields of human knowledge is represented in Wikipedia. However, various editor communities may have different perspectives on the boundaries of human knowledge needed to be incorporated into Wikipedia. For example, "the Association of Deletionist Wikipedians"[2] aims to restrict Wikipedia articles to strictly "encyclopedic" topics. Researchers have investigated comprehensiveness of Wikipedia in a variety of fields ranging from art, philosophy, and history to medicine, psychology, and science. This

---

[2] http://meta.wikimedia.org/wiki/Association_of_Deletionist_Wikipedians

stream of research captures the fields that are supposedly underrepresented or overrepresented in Wikipedia, and sometimes come up with complementary or conflicting results. Of course, Wikipedia's coverage of topics has grown over time. Thus, we arrange the articles in this section mainly chronologically, as well as by topics, because the results of earlier studies might not be representative of Wikipedia's current condition.

Multidisciplinary and general: Several, multidisciplinary and general examinations of Wikipedia coverage have compared the proportional growth of fields of knowledge on Wikipedia. By sampling 3,000 articles from the 2006 English Wikipedia and categorizing them against the Library of Congress categories, Halavais and Lackaff (2008) found categories such as social sciences, philosophy, medicine and law underrepresented in Wikipedia compared to statistics from Books in Print. The two latter categories, however, had on average comparably large article sizes. They identified science, music, naval studies and geography as overrepresented, with music probably benefiting from fan contributions and other categories from the mass-insertion of material from public data sources. For instance, much basic geographical information of U.S. cities and towns is inserted from the 2000 United States Census. In addition, the names of ships in the U.S. and British fleets and details of weapons have been readily obtained from public sources in order to create stub articles. When compared to three specialized encyclopedias in linguistics, poetry and physics, they found many expected articles to be missing. Halavais and Lackaff also noted some peculiarities in Wikipedia, such as extensive list of arms in the military category, comic fans to some extent driving the creation of articles in the fine art category, and voluminous commentary on the Harry Potter series in the literature category. We note that the idea of "underrepresentation" is a legitimate concept, as any general encyclopedia should adequately cover any important field of knowledge. However, we consider the concept of "overrepresentation" rather dubious, that is, the idea that Wikipedia had too many articles on certain topics, such as video game characters or specific TV show episodes. Traditional constraints on topic coverage arise from past ages with stricter resource limitations in printed pages and human contributors. When these constraints are lifted, as Wikipedia does, then there is little justification to restrict any breadth to coverage on topics that manifest a strong interest among willing contributors. We note, though, that not all Wikipedia contributors agree with this point, and some try to restrict it to the traditional encyclopedic topics.

In another multidisciplinary attempt, on a cross-section of 446 articles randomly picked from *Encyclopædia Britannica,* Wikipedia articles lacked entries for 15, e.g., "Bushman's carnival," "Samarkand rug" and "Catherine East" (Wedemeyer et al., 2008). All 192 random geographical articles picked from *Britannica* had corresponding articles in Wikipedia. Of 800 core scientific topics selected from biochemistry and cell biology text books, 799 could be found in Wikipedia. Wedemeyer et al. concluded that science is better covered than general topics and that Wikipedia covers nearly all encyclopedic topics.

West and Williamson (2009) randomly chose 106 Wikipedia articles from various topics, and checked for the internal completeness of each article, referring to the depth of each article and the amount of details included. With an average of 4.2 out of potential 7, it marginally passed their required criteria for completeness of articles; completeness appeared to be the weakest aspect of Wikipedia content quality for their sample, though they believed that their account of completeness was limited because they did not examine outward links.

Examining the implications of gender imbalance in Wikipedia community, Lam et al. (2011) demonstrated that certain broad areas of Wikipedia content involved a significantly higher percentage of female contributions, and this affects the internal comprehensiveness of articles in such topics; article length is used as a proxy for comprehensiveness. The "art" and "people" categories are the two topics to which female Wikipedians were more interested to contribute than were males; consequently these were less comprehensive and shorter in length. Science and geography, in contrast, were topics with higher male than female contributor percentages, and are better covered. Reagle and Rhue (2011) likewise argued that the gender imbalance in the Wikipedia community has resulted in gender bias in content.

Comparing the biographies of individuals in Wikipedia and Britannica, they found that although Wikipedia had a higher number of women biographies, the proportion of missing biographies for women to that of men was relatively higher in Wikipedia than in Britannica.

Kittur et al. (2009) developed an algorithm that would assign a topic distribution over the top-level categories to each Wikipedia article. After evaluating the algorithm on a human-labeled dataset, they examined the English Wikipedia and found that "Culture and the arts" and "People and self" to be the most represented categories. Between 2006 and 2008, they found that "Natural and physical sciences" and "Culture and the arts" categories grew the most.

Royal and Kapila (2009) compared the number of words in sets of the Wikipedia articles for years from 1900 to 2008 (e.g., http://en.wikipedia.org/wiki/1900). They found that articles for recent years tended to be longer, that is, recency could somewhat predict coverage. The results were not homogeneous across all types of articles, as the correlations varied across articles associated with Academy award-winning films, "artist with #1 song", and Time's person of the year. Moreover, in their comparison with 100 articles from the Micropædia of the *Encyclopædia Britannica*, they found that 14 of these had no Wikipedia entry.

Medicine and health: Medical and health topics on Wikipedia are among the most investigated across the comprehensiveness studies. In 2005, an early examination of Wikipedia's content coverage reported insufficient representation of medical informatics on Wikipedia, with many important topics missing (Altmann, 2005). In 2008, Clauson et al. (2008) compared medical drug information on Wikipedia and Medscape Drug Reference (MDR), a free online traditionally edited database. They found that Wikipedia could answer fewer drug information questions, e.g., about dosage, contraindications and administration. In the evaluated sample, Wikipedia had no factual errors but a higher rate of omissions compared to MDR. The authors could also find a marked improvement in the entries of Wikipedia over a just 90 days period. The study went on to mainstream media with headlines such as "Wikipedia often omits important drug information" (Harding, 2008) and even "Why Wikipedia Is Wrong When It Comes To Prescription Medicine" (CityNews.ca, 2008). However, as noted by some Wikipedians[3], the study neglected the fact that one of the Wikipedia manuals of style explicitly requests: "Do not include dose and titration information except when they are notable or necessary for the discussion in the article." Thus, in one of the eight examined question categories in Clauson et al.'s study, the omissions were quite possibly intentional.

Health topics on Wikipedia have also been examined in terms of accessibility through search engines on the Web. Using search engine optimization techniques, Laurent and Vickers (2009) investigated the Google ranking of the English Wikipedia for health topics. The queries were 1726 keywords from an index of the American MedlinePlus, 966 keywords from a NHS Direct Online index and 1173 keywords from an American index of rare diseases (U.S. National Organization of Rare Diseases). They compared Wikipedia to .gov domains, MedlinePlus, Medscape, NHS Direct Online and a number of other domains. They found the English Wikipedia ranked among the first ten results in 71– 85% of search engines and keywords tested, concluding that the English Wikipedia is an outstanding source in comparison to the other sources providing online health information.

Wikipedia seems to be rich enough about disease information to be useful as learning material. Kim et al. (2010) examined the usefulness of Wikipedia content in covering the pathology informatics educational curriculum, and found that it covers 90% of the curriculum with high-quality, comprehensive and current articles beneficial for both beginning and advanced learners; however, about half the articles were tagged as needing improvements such as more citation. In addition, Leithner et al. (2010) investigated the quality of Wikipedia information on osteosarcoma, a type of cancer, in three aspects of scope, completeness and accuracy. Three independent observers scored the answers to twenty questions. They judged that the

---

[3] http://en.wikipedia.org/wiki/Wikipedia_talk:WikiProject_Medicine#Drug_Information_in_Wikipedia

information provided on English Wikipedia is good in terms of quality, but still inferior in comparison to that provided by professional health websites like the US National Cancer Institute (NCI). Thus, they suggested maintaining a high quality for Wikipedia articles by inserting external links to these professional sources.

History: Rosenzweig (2006) compared Wikipedia's history articles with those of other online and offline sources, concluding that Wikipedia "beats Encarta but not American National Biography Online in coverage" of the topics, as well as in the amount of detailed description of the topic in each article (2006, p. 129). Rector (2008) examined the comprehensiveness and accuracy of Wikipedia articles in comparison to three other major reference resources: Encyclopaedia Britannica, The Dictionary of American History and American National Biography Online. She extensively analyzed the content of nine history articles chosen randomly from purposefully selected groups of places, events, biographies, and movements/phenomena. She concluded that while Wikipedia is an appropriate model for peer-production of encyclopedic reference material, it is not as robust as others in terms of accuracy and comprehensiveness of articles.

Psychology: Schweitzer (2008) examined the coverage of psychology-related topics on Wikipedia, and reported that not only were they well covered, but they also displayed on top of the major search engines. Students were found to use Wikipedia for personal and school-related activities, but generally not as academic references. They used a list of 100 commonly used and most important psychological concepts, and found that more than 80% of the concepts are substantially covered in Wikipedia in detail, and another 10% received some brief coverage in more general articles. They describe Wikipedia's coverage of psychological articles as "impressively comprehensive" (p.84).

Philosophy: Philosophy appears to be well-represented on Wikipedia in terms of biographies of philosophers, but perhaps not enough in terms of philosophical ideas and arguments. For twentieth-century philosophers, Elvebakk (2008) compared Wikipedia against two online peer-reviewed resources, The Stanford Encyclopaedia of Philosophy and the Internet Encyclopedia of Philosophy, with respect to coverage of gender, nationality and discipline. She concluded that Wikipedia in 2008 represented philosophy topics essentially the same way as more traditional resources. Wikipedia had far more articles about the philosophers than the two other resources and only some minor differences in proportions, such as a smaller proportion of German and French philosophers. Similarly, Bragues (2009) tested "the quality of Wikipedia, [by] sampling ... articles relating to seven top Western philosophers" (p.117). However, he found out that on "average, the online encyclopedia captured 51% of the expert consensus surrounding the seven philosophers examined" (p.151). All of the analyzed philosophers' pages had a strong biography section, "arguably too strong" (p.152). "This could reflect the fact that contributors to Wikipedia's philosophy pages have less experience and confidence grappling with philosophical analysis. It may be that, compared to academic philosophers, Wikipedians on average find it less pleasurable to engage philosophic arguments and prefer to focus on the characters and histories of famous personages" (p.152). Bragues concluded that he "was unable to uncover any outright errors" and that the "sins of Wikipedia are more of omission than commission" (p.152).

Communication: Communication maybe one of the few areas reported to be weakly represented on Wikipedia. To analyze the public impact of communication research, Rush and Tracy (2010) argued for measuring the Wikipedia presence of an academic field as a proxy for the public impact of the field, as presence and accessibility are the necessary conditions for having impact. They thus concluded that communication research did not have the impact it was supposed to, and offered suggestions to improve this situation.

Biology: The studies that examined Wikipedia's coverage of biological sciences provided rather positive evaluations. Jancarik and Jancarikova (2010) examined the appropriateness of Czech Wikipedia material for preparing and teaching biology and mathematics courses. They demonstrated that the English Wikipedia properly covered the topics with highly detailed articles, but the Czech Wikipedia, whose

scientific topics are mostly translated from the English version, included less detail and covered fewer topics, which made it inappropriate to use in an e-learning course. Atanassova (2011) looked into how bioengineering topics were covered in Wikipedia. The study identified many Wikipedia article categories, projects and portals related to bioengineering topics.

Other specific topics: Besides these main fields studied in terms of Wikipedia coverage, multiple other areas have been examined. Forestry topics on Wikipedia have been studied for quality and extensiveness. Since they found the related articles originally very limited in 2010, Radtke and Munsell (2010) assigned students to create and improve forestry articles, and were pleased to find that even after the student assignment, numerous Wikipedia contributors actively continued to develop the articles. In another study, chemistry topics were compared on German Wikipedia to the chemistry encyclopedia Rompp Online based on 30 articles on chemical thermodynamics (Korosec, Limacher, Lüthi, & Brändle, 2010). After evaluating various aspects of the content quality of the two references, they found that German Wikipedia articles were more complete and lengthy than their counterparts in Rompp Online. Comparing Wikipedia as a thesaurus and the agriculture-specific thesaurus of Agrovoc, (Milne, Medelyan, & Witten, 2006) demonstrated that Wikipedia adequately covered a substantial proportion of agricultural concepts and their semantic relations; each concept is represented as an article in Wikipedia, and the relations are depicted as links between them. Political science is another area examined for comprehensiveness on Wikipedia, and found to suffer from "extremely frequent" omissions; nonetheless, the existing content was found to be "almost always" accurate (Brown, 2011). The older the topic of Wikipedia articles, the more the omissions that were found.

Some researchers have studied methodological issues in measuring the comprehensiveness of Wikipedia articles. Perception-based measures of completeness of Wikipedia articles have been shown to be moderately reliable, while this is not necessarily the case for other aspects of information quality; such differences are due to the nature of these various aspects of quality (Arazy & Kopak, 2011). Stvilia et al. (2007) suggested measuring intrinsic completeness of Wikipedia articles based on the number of internal links, broken internal links, and article length. They demonstrated that such measures successfully discriminated between lower and higher degrees of completeness.

Overall, in almost every domain of knowledge examined, Wikipedia is found to be one of the most comprehensive sources in existence with an extremely broad range of coverage. Even in the few fields where it was found to be not as broad in coverage as comparable resources, such as communications and forestry, the extent of coverage might broaden rapidly in just a few short years, as some studies found (Kittur et al., 2009; Royal & Kapila, 2009). Indeed, Wikipedia's open approach to contributor inclusion guarantees that with the passage of time, its comprehensiveness could only increase. However, the passage of time is not only a factor for letting Wikipedia catch up with its shortcomings; it also involves the creation or discovery of new knowledge that the encyclopedia would need to capture to remain usefully comprehensive. Thus, the next aspect of quality we examine involves Wikipedia's currency, investigating how well it maintains a record of the most recent state of human knowledge.

Currency

Currency refers to the degree to which Wikipedia articles reflect up-to-date information about their topics. Currency is considered an essential component of article quality. Wikipedia's live, continuous online publishing model has generally proven a major strength in comparison to other encyclopedias and information resources, both online and offline.

Stvilia et al. (2007) measured the currency of Wikipedia articles in general by the number of days from the last update of each article. They demonstrated that the higher-quality articles (median currency of 3 days) were significantly more current than the lower-quality ones (median 46 days).

Most studies of Wikipedia's currency have restricted their scope to specific knowledge domains. In a study on twentieth century philosophers, Wikipedia had far more articles on philosophers born after the

Second World War than two other online encyclopedias, The Stanford Encyclopedia of Philosophy and The Internet Encyclopedia of Philosophy (Elvebakk, 2008). Laurent and Vickers (2009) demonstrated that Wikipedia topics on health information were getting updated expeditiously by new events and findings announced in news. Kim et al. (2010) recognized currency as the strongest aspect of Wikipedia articles, with an average of 112 revisions per article over one year in a sample of pathology informatics topics. They concluded that the more the article was of general interest to the Wikipedia community, the higher the number of edits and revisions was.

Lack of currency may even harm the other aspects of article quality like reliability. In a comparison between Wikipedia and Medscape, Clauson et al. (2008) found four factual errors in Medscape among 80 articles examined. Two of these occurred due to lack of timely updates. In contrast, they found no factual errors in Wikipedia.

Although Wikipedia is often current, the fact that large bodies of work available from the public domain (which are often many decades old) are sometimes imported en masse compromises the currency of certain parts of Wikipedia content. For instance, the Danish Wikipedia has a large number of articles copied more or less unedited from two old reference works with expired copyrights: *Dansk biografisk Leksikon* and *Salmonsens Konversationsleksikon*. The age of the works affects the language and viewpoint of the Wikipedia articles (Bekker-Nielsen, 2011). Such risks might also occur in the English Wikipedia, where many articles feature imports from the 1911 edition of *Encyclopædia Britannica*. However, such importation of old material has not been so substantial as to degrade the currency of Wikipedia as a whole. On the contrary, even with such risks, Wikipedia has nonetheless been found to be generally much more up-to-date than the present-day *Britannica* (Wedemeyer et al., 2008).

In every study comparing Wikipedia's currency with that of other comparable reference sources, Wikipedia has been uncontested in its ability to update its information with current knowledge. Even in the case of a possible systematic compromise of currency, that is, in importing content from very old sources, Wikipedia generally remains nonetheless more current than the comparable contemporary reference. These findings are a strong testament to the advantages of the wiki approach in enhancing this important aspect of quality.

Readability and Style

Readability is an important aspect of the quality of an encyclopedia article. It is quite different from other quality measures, since it is completely distinct from the accuracy or usefulness of the articles; it has to do with the accessibility of the presented article to readers. Without a readable writing style, readers would have difficulty reading and benefiting from the articles, regardless how high quality they might be according to other quality standards.

Stvilia et al. (2007) defined readability in terms of the complexity of an information object that is "the degree of cognitive complexity of an information object relative to a particular activity". They operationalized complexity using the Flesch and Kincaid readability scores, and demonstrated that median readability score of Wikipedia featured articles was significantly better than that of other random articles, though it was not yet sufficient. Having an average Flesch score of 36 and 27 respectively for featured and random articles, Wikipedia seemed to be easily readable for university graduates, but not for younger and less literate people that need at least a Flesch score of 60 or higher to be easily readable.

Some studies examined Wikipedia's readability in its own right, without making external comparisons. Positively, these studies found many well-written articles. Negatively, the same studies found many poorly-written articles, and found that the overall quality is rather inconsistent across the encyclopedia. Dalby (2007) commented generally on the language versions of Wikipedia, focusing mainly on the English Wikipedia. He noted that the quality of English is very inconsistent, and that many non-native English speakers contribute, leading to poor quality writing in some articles, especially those on international topics. He noted that while about 80% of English Wikipedia editors are from English-

speaking countries, the other 20% are not. West and Williamson (2009) investigated the quality of Wikipedia articles, and found that Wikipedia articles are objective, clearly presented, reasonably accurate, and complete. However there is little consistency across articles, and there are some poorly written articles containing unsubstantiated information and providing shallow coverage of their topics. Clark et al. (2009) found Wikipedia articles can be distinguished not only by their various topics, but also by their structural form, i.e., writing genre. The genre may evolve as editors extend and change the articles.

Ehmann et al. (2008) employed various operationalizations of readability to compare Wikipedia articles in the humanities, hard sciences, and soft sciences; such measures included Flesch–Kincaid score, Flesch score, number of passive sentences, number of words per sentence, and number of sentences per paragraph. While Wikipedia articles in general appeared to be low in readability, hard science articles were much easier to read than humanities and soft science articles. In contrast to such objective quantitative measures of readability, more subjective perception-based measures of readability and style have been found to be marginally reliable, because of generally low consistency on such quality aspects across evaluators (Arazy & Kopak, 2011).

Purdy (2009) conducted an in-depth scholarly study of writing composition characteristics in three Wikipedia articles. He argued that Wikipedia represents an important form of writing today—online collaboration. He observed that although Wikipedia represents a new form of dynamic, unstable knowledge, it nonetheless manifests the traditional writing composition elements of revision, collaboration and authority. *Revision* refers to the free editing and history capabilities of Wikipedia through which anyone can monitor all the previous versions of any article and edit the last version. *Collaborative* work is an inherent part of every textual production, and it is supported in Wikipedia through talk pages attached to each article. It supports the process of developing and revising textual knowledge. In any knowledge work, *authority* is demonstrated by referencing to verifiable knowledge sources. Wikipedia supports authority by encouraging editors to put references for every fact they add. References in Wikipedia not only support the claims made in the text, but also provide sources for further reading.

Some studies compared Wikipedia's readability with that of other comparable online resources. These studies varied in their results: some found Wikipedia articles generally equally readable, some less so, and others found Wikipedia generally more readable. Elia (2009) compared the readability and maturity of Wikipedia articles with those of the Britannica Online encyclopedia in terms of varied readability indexes like word length, sentence length, and lexical density; she found no significant differences in these quantitative measures. Korosec et al. (2010) compared student use of the German Wikipedia and the chemistry encyclopedia Rompp Online in the area of chemical thermodynamics. They found that while students use both, Rompp Online is victim of its exactness and academic writing style. Wikipedia is more comprehensive and more easily readable, two characteristics that are very important to students. They concluded that while both resources are good for initiating research, students should learn how to use both peer-reviewed and non-peer-reviewed material in their learning. Comparing Wikipedia's cancer information from August 2009 with the US National Cancer Institute's Physician Data Query (PDQ), Rajagopalan et al. (2010, 2011) found that Wikipedia had lower readability as measured by the Flesch–Kincaid readability test.

Emigh and Herring (2005) examined the genre of collaborative authoring. They measured the degree of formality and standardization as an essential aspect of writing style, and compare four sources: Wikipedia articles, Wikipedia discussion pages, Everything2, and the Columbia Encyclopedia. Whereas Everything2 is an online collaborative general encyclopedia very much like Wikipedia, Colombia is a traditional printed one. Surprisingly, Wikipedia is not distinguishable from Colombia in terms of formality; both appeared to be more formal than Everything2. Wikipedia discussion pages were less formal than the other three sources. The authors ascribed Wikipedia's formality and standardization, in comparison to Everything2, to the post-production editorial controls that encourage people to use a harmonized and standard language style. In line with such a focus on editorial processes, Den Besten and Dalle (2008)

studied the Simple English Wikipedia, a distinct Wikipedia (that is, with completely separate articles from the regular English Wikipedia) that limits its vocabulary sense and grammatical structure to facilitate reading by children and by learners of English. They investigated the editorial processes and found that the tagging system encouraging editors to tag non-simple articles has been successful. However, this success has diminished as the number of articles has increased, because the editorial resources are limited and the system does not allow identification of all non-simple articles and follow-up to see if the article is improved. They suggested editorial companions and bots that facilitate monitoring of article readability scores.

Overall, the studies that have examined the writing style and readability of Wikipedia articles have generally found that they are at least as easy to read as their online and offline counterparts (Elia, 2009); however, results were varied when specific knowledge domains were investigated (Korosec et al., 2010; Rajagopalan et al., 2010, 2011). Wikipedia's writing style was found to be inconsistent (West & Williamson, 2009), especially concerning international topics (Dalby, 2007). We believe that Wikipedia's success and shortcomings in readability are both due to the mass collaborative open editing policies (Purdy, 2009). On one hand, because many people over time read and correct articles, Linus' law applies: "Given enough eyeballs, all bugs are shallow." That is, because many people have the opportunity to see readability problems and are empowered to correct them, even anonymously, articles should tend to gravitate towards increased readability over time (Duguid, 2006). On the other hand, because Wikipedia's culture eschews any kind of central editorial control, it is impossible to maintain a consistent writing style, and so many less-read articles might languish with a messy writing style for years (West & Williamson, 2009).

Reliability

When people talk about Wikipedia's "quality", it is most likely that reliability is the specific quality dimension that is most often meant. Thus, reliability has always been one of the main concerns of Wikipedia users and is one of the most widely investigated aspects of the Wikipedia phenomenon. This quality dimension is variously called reliability, accuracy, and freedom from errors. Here we treat studies where subject experts empirically evaluated the reliability of Wikipedia articles. In restricting our treatment to evaluations by subject experts, we consider these experts to be the most objective means of evaluating the degree to which the contents of a Wikipedia article corresponds to the true state of accepted knowledge.

We have a very narrow definition of Wikipedia reliability studies, and so we need to clarify how and why we distinguish our focus from many similar yet distinct kinds of treatment. We distinguish reliability from measures of trustworthiness, which we generally call "credibility" — Contributor Perceptions of Credibility and Reader Perceptions of Credibility refer to contributor and user perceptions of reliability, respectively— and are discussed in other publications of our review project (Okoli et al., 2014, sec. Reader Perceptions of Credibility, 2012, sec. Contributor Perceptions of Credibility). These measures of credibility are subjective and perception-based, as distinct from the more objective measures that we call reliability. Indeed, subjective measures of credibility have been found to be less statistically reliable (Arazy & Kopak, 2011), meaning that scores by different evaluators of the same articles vary widely. We also distinguish our treatment here from Computational Estimations of Trustworthiness (Okoli et al., 2014), which uses computational methods to estimate how much credence a reader ought to lend an article. Concerning experts being able to determine "truth", we also distinguish these studies from the large body of research that has examined epistemological questions of how Wikipedia reflects the controversial notion of "truth"; we discuss such studies in another review (Okoli et al., 2012, sec. Epistemology). Finally, although many studies have used students to evaluate the reliability of Wikipedia articles, by definition students are not experts, and so we consider such studies to be Reader Perceptions of Credibility (Okoli et al., 2014).

We group this body of work into three subsections. Many studies examined the reliability of articles either on their own, or in comparison with other reference sources. A second group of studies examined the quality of citations from Wikipedia articles to external sources. A third group examined trends in the reliability of articles over time.

With continuous revision, the reliability of Wikipedia articles has generally improved over time (at least, for existing articles; new articles start from ground zero in terms of quality). Thus, we arrange the articles in this section mainly chronologically, since the results of earlier studies might no longer accurately represent Wikipedia's most recent condition.

Table 2. Reliability Assessments of Wikipedia

	Reference	Domain	Comparisons against Wikipedia	Sample	Evaluators
Positive or Equivalent Evaluations	Giles (2005)	Natural sciences	Britannica	42 articles	Blinded experts
	Chesney (2006)	Various	N/A	30 articles	Academic Experts
	Kim et al. (2010)	Pathology informatics	N/A	40 articles	Authors
	Devgan et al. (2007)	Surgical procedures	N/A	35 articles	Experts
	Magnus (2008)	Philosophy	N/A	36 fibs	Author
	West and Williamson (2009)	Various	N/A	106 articles	Authors
	Rosenzweig (2006)	U.S. history	Encarta & American National Biography Online	25 articles	Author
	Pender et al. (2008)	Health	UpToDate and eMedicine and AccessMedicine	3 articles	Blinded experts
	Rajagopalan et al. (2010, 2011)	Cancer information	US National Cancer Institute's Physician Data Query (PDQ)	10 articles	Medically trained personnel
Negative or Inferior Evaluations	Mercer (2007)	mental health	N/A	4 articles	Author
	Rector (2008)	History	Britannica & The Dictionary of American History & American National Biography Online	9 articles	Author
	Clauson et al. (2008)	Drug information	Medscape Drug Reference (MDR)	80 Questions	Authors
	Leithner et al. (2010)	Osteosarcoma	NCI patient and professional site	20 Questions	3 independent observers
	Brown (2011)	Political science	N/A	Thousands of articles	Author
	Lavsa et al. (2011)	drug information	drug package information and certain authoritative databases	20 articles	Residency-trained pharmacists

Reliability Assessment of Wikipedia

Some of the most popular Wikipedia studies—that is, those that have received the most press attention—are those that face off Wikipedia against authoritative sources of information to compare their respective reliabilities. Although very many such comparisons have been conducted, here we discuss only the scholarly ones, as we explained in the introduction. Results have been mixed, with some studies evaluating Wikipedia quite favorably, and others not as much. We group these studies accordingly, as depicted in Table 2.

Positive or Equivalent Evaluations: Some empirical studies have found Wikipedia at least equal in reliability to well-established reputable sources. Some have found Wikipedia even superior.

The most famous scholarly assessment of Wikipedia is a comparison of selected science articles in Wikipedia and *Encyclopædia Britannica* conducted by *Nature*, the leading science journal (Giles, 2005). Giles found Wikipedia's accuracy comparable to those of *Britannica*. The articles were masked as to their provenance and evaluated by scientists who had published in *Nature*. The scientists found that among 42 articles, Wikipedia contained more factual errors, omissions and misleading statements (162, with an average of 4 per article). However, *Britannica* was not far behind (123 errors, average 3 per article). Both encyclopedias contained "serious errors, such as misinterpretations of important concepts" (2005, p. 900). Although *Britannica* fared better on this examination, its finding was shocking and was widely considered a major blow against *Britannica* and a boon to Wikipedia, considering that Wikipedia was only four year old at the time, compared to *Britannica* at 232 years old in 2005. The study was not itself peer-reviewed, and was vociferously contested by *Encyclopædia Britannica*;[4] however, *Nature* defended its analysis.[5]

There are studies analyzing content accuracy of Wikipedia by means of absolute measures using expert opinion but not in comparison to other sources. Among them, Chesney (2006) identified Wikipedia articles as highly credible. Kim et al. (2010) assessed the pathology informatics topics on Wikipedia in terms of comprehensiveness, quality, and currency. They found that the examined articles are of good quality with few errors; they judged that the articles can be used in a course curriculum for teaching to beginner and advanced learners. Devgan et al. (2007) surveyed medical doctors concerning 39 common surgical procedures. They could find 35 corresponding Wikipedia articles, with all of them judged to be without overt errors. The researchers could recommend 30 of the articles for patients (22 without reservation), but also found that 13 articles omitted risks associated with the surgical procedure. Analyzing accuracy of a varied sample of 106 randomly selected articles, West and Williamson (2009) found Wikipedia content as having "reasonable accuracy" with a score of 5.1 out of 7. Magnus (2008) made an experiment of adding purposely faulty information to Wikipedia article. He found that about "one third to one half of the fibs were corrected within 48 hours."

Although he did not conduct an empirical investigation, Fallis (2008) analyzed Wikipedia's reliability from an epistemological perspective. He argued that it is most meaningful to judge Wikipedia's reliability relative to that of other encyclopedias such as *Britannica*, rather than on some dubiously absolute standard. On that basis, he considered Wikipedia a superior contemporary source of knowledge.

When compared to some other resources, Wikipedia was found to be accurate in reporting names, dates, and events in U.S. history; in 25 biographies only four clear-cut factual errors, mostly small and inconsequential, were found: "Wikipedia ... roughly matches Encarta in accuracy" (Rosenzweig, 2006, p. 129). Pender et al. (2008) compared Wikipedia with UpToDate and eMedicine; they found roughly the same level of factual errors in these three sources. However, another source they compared, AccessMedicine, contained no factual errors in the three articles examined. Rajagopalan et al.

---

[4] http://corporate.britannica.com/britannica_nature_response.pdf

[5] http://www.nature.com/nature/britannica/index.html

(Rajagopalan et al., 2010, 2011) examined Wikipedia's cancer information in August 2009 against that in the US National Cancer Institute's Physician Data Query (PDQ). They found that Wikipedia had similar accuracy and depth compared to the professionally-edited resource.

Negative or Inferior Evaluations: Some empirical studies have found Wikipedia of inferior quality to well-established reputable sources. Some have concluded that Wikipedia is of such poor quality that it is inadvisable to use it; these more strongly negative recommendations tend to accompany unfavorable evaluations in healthcare topics.

Mercer (2007) reviewed some key mental health topics in Wikipedia and found them generally lacking in quality, mainly because of what he perceived to be the influence of contributors lacking genuine professional expertise on the subjects. However, he recognized Wikipedia's importance and potential and recommended a number of measures that could hopefully improve the quality of articles. Unfortunately, most of these recommendations involved contributors revealing their real-world identities, which conflicts with Wikipedia's strong policy of permitting anonymous participation and emphasizing quality of content over the qualifications of contributors.

Cautionary notes have been made for the open-wiki model in cases where potentially hazardous procedures are described (Caddick, 2006). In particular, medical procedures and pharmaceutical compounds may call for complete and accurate description. Clauson et al. (2008) compared Wikipedia and Medscape Drug Reference (MDR), a free online "traditionally edited" database, for medical drug information. They found that Wikipedia could answer fewer drug information questions, e.g., about dosage, contraindications and administration. In the evaluated sample, Wikipedia had no factual errors but had a higher rate of omissions compared to MDR. Moreover, Clauson et al. found a marked improvement in the entries of Wikipedia over a just 90 days period. Very similar results were reported about political science articles of Wikipedia. Investigating accuracy and completeness of political science articles, Brown (2011)examined some objective information on thousands of articles related to elections, candidates, and office holders. He found that Wikipedia articles were "almost always" accurate, but suffered from "extremely frequent" omissions, making them less attractive as an information source.

Leithner et al. (2010) investigated the scope, completeness, and accuracy of information for osteosarcoma on English Wikipedia in April 2009, compared with patient and professional sites of the US National Cancer Institute (NCI). Although they found Wikipedia's information to be generally good, it scored lower compared to the two NCI versions (though this was statistically significant only for the professional version). Thus, they suggested adding external links to these websites on Wikipedia articles.

Lavsa et al. (2011) compared the drug information for twenty of the most frequently prescribed drugs in the United States with the drug package information and certain authoritative databases. They found that the Wikipedia articles were all incomplete in providing full drug information, often missed important details, and were often inaccurate. They recommended against its use by pharmacology students for drug information. As mentioned earlier, part of the missing information in health-related articles could be quite intentional, as Wikipedia policy advises not to include drug dosage information in articles.

In the Comprehensiveness section of this review, we discuss the only non-medical study that concluded with a negative evaluation of Wikipedia. Rector (2008) examined the comprehensiveness and accuracy of history-related articles in comparison to three other reference resources. Although she considered Wikipedia a proper model for collaborative development of reference materials, she judged it inferior to the comparator resources.

Beside all these positive and negative evaluations of Wikipedia reliability that we have discussed, Magnus (2009) argued against what he considered the simplistic perspective of whether Wikipedia is reliable or not; rather, "interacting with Wikipedia involves assessing where it is likely to be reliable and where not".

It would be overly simplistic to merely count the number of studies that considered Wikipedia generally reliable against the number that did not (for example, eight versus five in Table 2). Rather, it is more meaningful to note that it is predominantly (though not uniformly) the health-related studies that were more cynical of Wikipedia's reliability—this can be expected considering the life-staking nature of health information. However, it is noteworthy that reliability studies in almost all other domains (including a couple in health) concluded that Wikipedia is generally a reliable source of information.

Verifiability: Citing Other Sources

Verifiability is one of the main community standards governing quality contribution to Wikipedia articles, which requires Wikipedia editors to make their contributions to Wikipedia verifiable by supporting them with trustworthy external sources and citations. Some studies have examined verifiability and quality of citation sources as important proxy measurements of Wikipedia's reliability. Citing reliable sources within Wikipedia articles is not only an indicator of the reliability of the content, but also a way of guiding readers to access further information in other sources. Since Wikipedia's quality could be expected to evolve over time, we present these findings chronologically; the most recent studies might be more reflective of Wikipedia's current state.

Nielsen (2007) examined scientific citations in Wikipedia by examining the links from Wikipedia articles to the articles in scientific journals. He found "an increasing use of structured citation markup and good agreement with citation patterns seen in the scientific literature though with a slight tendency to cite articles in high-impact journals such as Nature and Science". Affirming such findings, Wedemeyer et al. (2008) reported that most well-developed articles had sufficient references comparable to a scientific review article, but some articles, even two featured ones, had insufficient referencing.

Assuming the open-access and peer-reviewed Stanford Encyclopedia of Philosophy to be a reliable source for articles of related topics, Willinsky (2008) studied how Wikipedia editors have drawn on this source to enhance the reliability and quality of articles. He demonstrated that Wikipedia had cited 80% of the entries in this scholarly encyclopedia. Moreover, most of the citations in Wikipedia articles led to academic journal and databases, and these citations were regularly used to access the original resources. Comparing nine history articles on Wikipedia and two other encyclopedias, Rector (2008) demonstrated that only 90% of the facts in Wikipedia were supported by verifiable resources, compared to 98% for Britannica and the Dictionary of American History. She thus judged Wikipedia articles to be less accurate.

There seems to be differences in verifiability across topics. Ehmann et al. (2008) reported that their examined sample of humanities articles included a significantly higher number of links to external resources than did soft and hard sciences articles. They found that Wikipedia articles had more internal links to other Wikipedia articles than external links to other sources. This is typical, however, of encyclopedia articles. Examining health-related Wikipedia articles, Haigh (2010) evaluated the quality of their source and supporting information. She found that health resources cited in Wikipedia articles were clearly identifiable and reputable, which made Wikipedia an appropriate resource for nursing students.

Huvila (2010) studied the less-investigated question of the original sources for content published in Wikipedia. He showed that "in spite of the popularity of online material a significant proportion of the original information is based on printed literature, personal expertise and other non-digital sources of information". He argued that this finding helps understand "how new Wikipedia articles emerge, how edits are motivated, where the information actually comes from and more generally, what kind of information may be expected to be found in Wikipedia".

In a critical evaluation of Wikipedia content verifiability, Luyt and Tan (2010) investigated the credibility of a randomly chosen set of Wikipedia articles about various countries' histories and found that most of the article contents were either not verifiable, or the resources were not credible ones like academic publications. They argued that the social context knowledge is more important in evaluating Wikipedia

than the typical focus on accuracy. They further argued that information literacy needs to be redefined to include this social aspect of knowledge construction, and that instructors should teach students about the disciplinary conventions of the context in which they write. This would create a new generation of better writers for Wikipedia articles, and for society in general.

Page (2010) compared Wikipedia with the Encyclopedia of Life, and argued that not only was Wikipedia stronger in search result rankings and contributor size, but also that it was especially valuable due to the potential direct linkages to other primary sources through DOIs or PubMed IDs. However, she found Wikipedia limited in taxonomic content.

In a comparison of Wikipedia with a peer-reviewed online encyclopedia, Scholarpedia, Stankus and Spiegel (2010a) evaluated how these two encyclopedias referenced books. They reported that although Wikipedia referenced books 40% less frequently, the books and authors referenced were as legitimate as those of Scholarpedia. However, Wikipedia more frequently referenced newer and more publicly accessible material and undergraduate books. Later, they extended their study beyond books to examine the journal citations in Wikipedia articles, and found very similar results (Stankus & Spiegel, 2010b). Wikipedia articles cited the journal publications fewer times, but those cited were more current in comparison to Scholarpedia. Moreover, the publications cited in Scholarpedia were most frequently from journals ranked in the top 20 of their respective fields; this was less the case for Wikipedia.

Generally, with few exceptions, the evaluations of the quality of citations and sources have been mostly positive. Although verifiability is a widely recognized criterion for evaluating encyclopedic reference sources, it is especially important to Wikipedia. In accordance with its Neutral Point of View policy, contributors are expected to support their contributions with proof of outside independent sourcing of the information to demonstrate that what they contribute is not their own opinion, but rather an externally verifiable fact or perspective. Thus, although the primary goal of Wikipedia's verifiability is to maintain quality, we suspect that its goal of neutrality is a strong driver for upholding the Verifiability policy.

Quality-Related Trends

Some studies observed the evolution of Wikipedia reliability measures over time, finding this to be mostly a progressive trend. Luyt et al. (2008) investigated how errors are spread out through the life of Wikipedia articles. They found that a significant number of erroneous edits occur in the earlier article edits, with 20% on the first day. Nielsen (2008) studied scientific citations in Wikipedia through time. He found an increasing use of structured citation markup, especially after mass insertion of gene and protein information and citations by a bot. Ortega (2009) quantitatively analyzed the eight-year trend across the top ten language editions of Wikipedia. He found that as of 2007, the number of new contributors had tapered off, as had the number of monthly contributions. Since most quality contributions are done by active and experienced users, he cautioned that the Wikipedia community needs to actively try to increase the number of active contributors, or else the quality of Wikipedia might suffer in future. However, he found that there was increasing activity in talk pages, which was related to increasing quality of their corresponding articles.

Antecedents of Quality

There has been great interest in understanding the factors that lead to the quality of Wikipedia articles. Characteristics of the group contributing to the article, patterns and processes of editing articles, and some others are among the main categories of such factors argued to be consequential for the quality of Wikipedia articles. Table 3 summarizes the antecedents of content quality in Wikipedia research.

Table 3. Antecedents of Wikipedia article quality

Factor types	Factors affecting article quality	Direction	References
Group Characteristics	Diversity/Heterogeneity	Positive	Carillo and Okoli (2011)
	Orientation towards content vs. Administration	Direct positive for content, and indirect positive for administration orientation	Arazy et al. (2011)
	Participation level	Controversial positive	Duguid (2006), Ehmann et al. (2008), Carillo and Okoli (2011)
	Member retention and turnover	Inverse U shape	Ransbotham and Kane (2011)
	Group size	Positive	Carillo and Okoli (2011)
	Shared experience	Positive	Carillo and Okoli (2011)
	Coordination	Directly and indirectly positive	Kittur and Kraut (2008), Stvilia et al (2008)
	Gender imbalance	Positive	Lam et al. (2011)
Editing Patterns and Processes	Contribution domain range	Negative	Adamic et al. (2010)
	Content editing vs. surface editing	Both equally positive	Jones (2008)
	Referencing practices	Positive	Rector (2008)
	Anonymity	Negative	Santana and Wood (2009), Anthony et al. (2009)
	Task conflict	Directly negative, and indirectly positive effect	Arazy et al. (2011)
Other Factors	Public attention	Positive	Rosenzweig (2006), Lih (2004)
	Public good	Positive	Rahman (2006, 2008)
	Free-riding and Free-editing	Positive	Rahman (2006, 2008)

Group Characteristics

In general, it is argued that some characteristics of the group who contribute to a Wikipedia article would affect the quality of the article content. The main group characteristics recognized to influence quality of articles includes group diversity, content orientation, size, contributor retention and turnover, and member activeness.

Arazy et al. (2011) studied how group composition aspects affect article quality on Wikipedia. The group members might be oriented towards either content or administrative activities. Content orientation is related to low levels of commitment and identification with the group, and relatively low participation, centered on just a few topics; administrative orientation is identified with high levels of group commitment and identity, and higher participation dispersed around various topics. The groups whose members are oriented towards content would produces higher quality articles compared to the ones

oriented towards administration activities. However, administrative orientation also enhances quality through reducing task conflict. Task conflict is when members of a group have varied understanding of the task to be done. Besides its direct negative effect on article quality, task conflict positively moderates the effect of cognitive diversity of the group on quality. In other words, diversity enhances Wikipedia article quality only in the presence of some task conflict; this is counterintuitive to our understanding of the task conflict function.

Focusing on the peer-production processes rather than its results, Duguid (2006) explained that the peer-production literature presents two laws responsible for quality of the product coming from software peer-production. First, the higher the level of member participation, the higher the quality of the results. Second, in whatever people produce, good quality content remains and poor content goes away. However, he then argued that these two laws are not enough for knowledge peer-production systems like Wikipedia. Moreover, the direct positive relationship between participation and article quality was not always supported.

Ehmann et al. (2008) examined the relationship between article edits and the quality of articles on a sample of nine randomly selected Wikipedia articles, and demonstrated that contrary to the past findings, articles with highest number of edits were not necessarily of the highest quality.

Following such a process-oriented view, Carillo and Okoli (2011) investigated the group process mechanisms that contribute to the quality of articles in Wikipedia. Applying the Input-Process-Output approach and the Time, Interaction, and Performance Theory, the results provided evidence for the positive effects of group size and shared experience on both group process variables and group effectiveness; of group heterogeneity on group production; organizational support and member activeness on group well-being; member activeness on member support; and of organizational support and member activeness on group effectiveness.

Coordination of the group of editors is another process factor affecting article quality. This coordination may be explicit, by discussing how to write different parts of the article, or implicit, by few editors structuring the whole article beforehand. It is even argued that having a higher number of contributors would increase the quality of the article only if coordination exists between them (Kittur & Kraut, 2008). In line with this, Stvilia et al (2008) accessed Wikipedia's extensive discussion pages to observe the nature and directions of the conversations between the contributors, and how these collaborations operate to assure information quality in Wikipedia. They found Wikipedia particularly valuable in the richness of its textual data in documenting the process of discovering and correcting errors in the information. The lessons learned from the study of Wikipedia would be also applicable to other textual databases.

Besides all these dynamics of the contributing group, the retention and turnover rate of the group members influence the eventual promotion to and demotion from featured article status. Ransbotham and Kane (2011) found that it is optimal to have a mix of both new and long-term contributors to attain and maintain featured status. In other words, there is an inverse U shape relationship between member turnover and content quality: a moderate level of turnover would be optimal. This finding contradicted widely-held views that long-term retention of participants is an unmitigated good in online collaboration. Moreover, the balance between male and female contributors seems to be another contributing factor; it is negatively related to the comprehensiveness of Wikipedia content. Lam et al. (2011) argued that articles with higher proportion of female contributors (that is, having less gender imbalance) are likely to be shorter in length and less comprehensive.

Editing Patterns and Processes

Some studies have examined the patterns and processes of editing Wikipedia that have led to higher-quality articles. These refer to specific ways in which people edit and contribute to Wikipedia, including the scope and content of contribution domain, types of editing, and anonymity of editing processes.

Significant correlation is found between focus and quality of individual contribution across a range of traditional and modern knowledge sharing media including Wikipedia; the narrower the individual contribution domain is, the higher quality and even quantity it would have (Adamic et al., 2010).

Along with the scope of the editing pattern, the content of the patterns is at least equally important. Jones (2008) investigated the relationship between patterns of editing and article quality on Wikipedia. He found that revision histories of quality articles are populated equally by both content- and surface-related edits: content edits are concerned with providing the most accurate and relevant information about the topic, whereas surface edits improve the structure of the article and how the information is presented. The lower quality articles were populated by more content edits and fewer surface edits. Moreover, Wikipedia promotes editing patterns based on "the virtues of inclusion and accuracy at large scales" (Klemp & Forcehimes, 2010), and this is what shapes patterns of contribution to Wikipedia.

Furthermore, referencing verifiable resources enhances the quality of Wikipedia content. Rector (2008) compared nine Wikipedia history articles to their counterparts on two other encyclopedia including Britannica and the Dictionary of American History. She found that referencing verifiable resources was related to accuracy. The facts on Wikipedia articles were 90% verifiable with references and 80% accurate, while the other two encyclopedias had about 98% verifiability and 95% accuracy. Such results support the role that referencing verifiable resources play in enhancing accuracy.

The process of editing is as influential on content quality as is the pattern. The collaborative editing processes of Wikipedia offer the opportunity for deliberative democracy that encourages quality participation for anonymous contributors (Klemp & Forcehimes, 2010). Such collaborative editing enables capturing collective intelligence that could not be captured by traditional deliberation.

Anonymity is another aspect of the Wikipedia editing process viewed as having consequences for content quality. Santana and Wood (2009) criticized Wikipedia as socially irresponsible, because the anonymity of its contributors (screen names of registered users are still anonymous without further identifying information) makes Wikipedia content not so transparent as to its sources. They mainly argued from principles of ethical theory where opacity through anonymity is often used by powerful actors (such as world-leading websites) to mislead their clients. They illustrated their trepidation with an extended example of a very active contributor and administrator who misled the community by lying about his real-life credentials. Along similar lines, Anthony et al. (2009) argued that "registered participants, motivated by reputation and commitment to the Wikipedia community, make many contributions with high reliability."

These studies show that contribution domain and content- and surface-editing are important editing patterns influencing the quality of Wikipedia articles. In addition, anonymity is also an important characteristic of the editing process that might possibly compromise quality.

Other Factors

Other studies have identified that the attention a topic attracts and citations in the news are other important factors affecting the quality of Wikipedia articles.

Public attention to topics is major contributor to the quality of articles on Wikipedia. Rosenzweig (2006) argued that since people contribute to Wikipedia voluntarily and mostly on the topics of their interest, its content is more comprehensive concerning topics of interest to the general public. Accordingly, other topics with less widespread public interest remained less complete. Sometimes, such public attention is

attracted through article citation across other media. In one of the earliest academic studies of Wikipedia, Lih (2004) gave an introductory sketch of the then three-year-old endeavor. He analyzed Wikipedia articles cited in the news in a thirteen-month period to compare their quality before and after citation in the press. Although his quality measures were rudimentary computations based on the numbers of edits and of contributors, his analysis did show that increases in his quality measure were correlated with citation in the press (though he did not demonstrate that this increase was not due to the timeliness of the events rather than due to the press citation).

Ehmann et al. (2008) highlighted the first-mover effect through which the creator of an article has essential influence on how it is structured and developed. They demonstrated that significant amounts of the original text remain over time, and that this highly impacts the quality of the final article.

Wikipedia is unique in that it is a public good that people freely edit and contribute to. Rahman (2006, 2008) employed game theory and economic modeling to demonstrate that Wikipedia's uniqueness "as a public good, combined with free-riding and free-editing help to maintain the [large size and] reliability of Wikipedia" relative to other open source systems. In other words, being a public good allows the free-riding effect of not contributing and using others' contributions, being mitigated through the free-editing effect of permitting everyone to freely contribute.

Overall, the studies have found two general classes of antecedents that are important in contributing towards the quality of Wikipedia articles. Group characteristics, such as the number of contributors to an article, their shared experience, and contributor retention and turnover, are critical, since article quality depends on the people who make the contributions. Their editing patterns and processes are the other important class of antecedents, such as the domain of the contribution and the absence of conflicts in the editing task. Thus, these major aspects both related to the effectiveness of contributors as a group play important roles.

Featured Articles

Featured articles are those that have been carefully examined and approved by the community as being of all-round high quality. In the English Wikipedia, such articles are honored by being "featured" as the article of the day on the English Wikipedia main page. Most other language Wikipedias have a parallel concept of their highest quality class of articles, though they bear different names, such as "excellent article" (German, Danish, Norwegian), "articles in showcase" (Italian), "article of quality" (French), or "recommended article" (Finnish, Norwegian).

Many studies that want to investigate not just random Wikipedia articles, but only a selection of articles of reasonably high quality, have focused on featured articles. Thus, these studies have accepted the community's evaluation of these articles as high quality, and then used them to validate their proposed quality assessment method for Wikipedia articles.

A number of studies have used computational methods to detect or predict higher quality articles, and have used featured article status to operationalize and validate their computational quality measures. Dondio and Barrett (2007) developed a method for computationally differentiating featured articles from others by predicting trustworthiness of articles. Zeng et al. (2006) provided a method to predict trustworthiness of Wikipedia articles based on the revision history of the articles, and validated it using featured articles. They argued that Wikipedia was generally trustworthy, and that visualizations of article trustworthiness could enable users to access the more trustworthy versions of the articles and to avoid vandalism and malicious content. Blumenstock (2008) offered word count as a predictor of article quality. He argued that this simple metric was "considerably more accurate than the complex methods proposed in related work, and performs well independent of classification algorithm and parameters." (2008, p. 1096).

Stvilia et al. (2007) presented a context-independent framework for information quality (IQ) assessment. They tested their framework with two large datasets, one of which was English Wikipedia. In the

Wikipedia case, they used featured articles as benchmarks for article information quality. Featured articles "were valuable resources for designing and validating individual IQ metrics as well as for testing the entire IQ measurement model" (2007, p. 1732). They evaluated their model using 236 featured articles and 828 random articles. They conclude that the "IQ measurement model was shown to be successful in discriminating high-quality articles" (2007, p. 1732). In a similar study, McGuinness et al. (2006) measured trustworthiness of Wikipedia articles based on the ratio of the number of links from other articles to the specific article to the number of potential incoming links. The suggested link-ratio measure successfully differentiated featured articles from others as featured ones had a higher link-ratio.

Wikipedia's processes for promotion or demotion of featured articles appeared to be a key success factor for its quality. Viégas et al. (2007) recognized Wikipedia's featured article evaluation processes as a unique aspect of Wikipedia's success as it "blends elements of traditional workflow with peer production" (p.445). In a study on Italian Wikipedia, Monaci (2009) investigated the quality assessment processes of featured articles. She argued that the Wikipedia community quality assessment is different from that of others in that it assigns value-maintaining roles as Administrators, Burocrati and Checkusers who are devoted to daily management activities over and above the individual authorship.

However, the Wikipedia community's vetting process cannot be considered a sufficient assurance of quality. Lindsey (2010) examined the quality control procedures for Wikipedia featured articles, and found that 12 out of 22 investigated English Wikipedia featured articles actually do not follow Wikipedia's own criteria; thus, he concluded that the quality control procedures are ineffective.

Other studies have examined factors that led to such articles to be considered as high quality articles. Poderi (2009) investigated the relationship between revision patterns and quality of featured articles of Wikipedia. He examined two aspects of revision patterns: edits typology and the editors' distribution. His findings suggested that having a main author would increase the consistency and quality of Wikipedia articles. Moreover, major edits were not necessarily more important in enhancing content quality, but the minor edits played a major role in revision pattern of featured articles. Ortega (2009) quantitatively demonstrated that featured articles were mostly edited and developed by active and experienced contributors. Besides the intuition of the need for the experienced users, a mix of both new and long-term contributors also appeared to influence attaining and maintaining featured status (Ransbotham & Kane, 2011). Moreover, Jones (2008) investigated the relationship between patterns of editing and article quality on Wikipedia by comparing featured and non-featured articles. He found that revision history of quality articles are populated equally by both content- and surface-related edits; however, the lower quality articles were populated by more content edits and fewer surface edits.

On a similar line of inquiry, Stvilia and Gasser (2008) employed activity theory to analyze the pattern of change in information quality of featured articles. They found that the article structure and article revision patterns were influential on article quality, and that the presence of a principal author enhanced the quality of an article. Later, Stvilia et al. (2009) demonstrated that different Wikipedia communities had diverse understandings of quality, and employed various criteria for featured article identification. However, there are some edit-based metrics, such as the number of edits, which would be able to meaningfully explain featured article promotion/demotion across different Wikipedia communities. There is also further evidence supporting that not only the edit number, but also the number of distinct contributors would differentiate featured articles from others (Wilkinson & Huberman, 2007).

Goldspink (2009, 2010) studied the discussive editing styles of Wikipedians by comparing featured articles against controversial ones. In analyzing discussion pages related to both controversial and featured articles, he observed a meaningful correlation between the article group and the communication style, and concluded that the discussion style could explain the article status. He found that featured article discussion pages included more positive utterances and fewer negative ones in comparison to controversial articles. Despite the fact that mutual encouragement is clearly supported by Wikipedia recommended etiquette, it rarely happened in the communicative interactions that he observed. He

concluded that despite "a large number of rules, etiquettes and guidelines, explicit invocation of rules and/or use of wider social norms appeared to play a small role in regulating editor behavior" (p. 652).

Comparing featured articles against those nominated for such status but that did not achieve it, Carillo and Okoli (2011) examined how group process mechanisms have contributed to article status. The results provided evidence for the positive direct or indirect effects of group size, shared experience, group heterogeneity, and member activeness on group effectiveness in developing quality articles.

In general, studies of featured articles have been of two kinds. Many have focused on the featured articles as a particular phenomenon in the Wikipedia community. These have mainly investigated the characteristics that set featured articles apart from others, and have also examined the process that leads to an article reaching such a level of quality. Researchers of the other major kind of study have investigated Wikipedia articles in general, but have used featured articles as their samples in order to assure that they were working with relatively high-quality Wikipedia articles, rather than a purely random sample that could turn out some very poor or very new articles. In general, the studies have found that Wikipedia's featured articles are of fairly high quality, though some scholars felt that this is not always the case.

Size of Wikipedia

Wikipedia is one of the largest encyclopedias in the world, with articles on almost any topic. Many studies have measured its size and growth trends, with some attempting to explain these trends and investigating the micro- and macro-level factors enhancing or hindering Wikipedia's growth.

Growth Trends

Wikipedia size would be measured in terms of various elements such as articles, authors, edits and links, to provide a quantitative analysis of Wikipedia's growth and complexity. Using such varied measures, Voss (2005) showed that Wikipedia grows exponentially with varying growth rates among different languages. He highlighted using Wikipedia as a thesaurus since each article covers a single concept with links to related articles. Consistent with this, Smith et al. (2007) studied the network evolution of the English, German, and Japanese Wikipedias, and found that they go through "different acceleration regimes in their evolution" (p. 1).

In one of the earliest investigations of Wikipedia, Lih (2004) gave an introductory outline of the then three-year-old Wikipedia, and described it as "the largest form of participatory journalism to date" (p. 1); he demonstrated that the average size of English Wikipedia articles and the number of article edits had been steadily increasing during those years. Later, Ortega (2009) found that although the number of contributors and of monthly contributions on the top-ten language editions of Wikipedia by 2007 (when the study data was collected) was not increasing for two years, there had been an increase in discussion page size and activity during the same period that accompanied quality improvements.

The reasons behind Wikipedia's growth have been the target of many inquiries, and various explanations are provided. We may distinguish two levels of factors affecting Wikipedia size. Micro-level factors are those related directly to Wikipedia and its community, factors over which the Wikipedia community may have direct influence. Macro-level factors are those related to the higher-level external environment within which the Wikipedia community grows—the Wikipedia community would hardly be able to change or influence any of such features.

Micro-Level Size Factors

Links to non-existing articles seems to be one micro-level aspect of Wikipedia growth. Spinellis and Louridas (2008) demonstrated that most new articles are spurred by the creation of links to non-existing

articles added by other users; the constant ratio of such complete and incomplete articles assures self-sustaining and rapid growth of Wikipedia.

The "no space restriction" appears to be a contributing factor of Wikipedia's "long tail", referring to the numerous articles that treat relatively obscure subjects, sometimes at great length (Greenstein, 2007). Free editing and being a public good are other aspects of such growth. Rahman (2006, 2008) examined the reasons behind Wikipedia success in terms of size and quality. He explained Wikipedia's size based on "equilibrium contributions depending on the differences in types" (2006, pp. 91–92). Moreover, Wikipedia's uniqueness "as a public good, combined with free-riding and free-editing help to maintain the [large size and] reliability of Wikipedia" (Rahman, 2008, p. 96) relative to other open source systems.

Bots, that is, non-human agents, play meaningful roles in the steady growth of Wikipedia. Niederer and Van Dijck (2010) observed that "bots are systematically deployed to detect and revert vandalism, monitor certain articles and, if necessary, ban users, but they also play a substantial role in the creation and maintenance of entries" (2010, p. 1383). They pointed out that "human editors would never be able to keep up the online encyclopaedia if they were not assisted by a large number of software robots" (2010, p. 1377). However, their analysis of various Wikipedia language versions showed that the small language versions of Wikipedia were more dependent on bots than their larger counterparts; there was a somewhat inverse correlation between the percentage of bot edits and the size of a Wikipedia language version. For example, the German Wikipedia "has only 9 per cent bot activity" (2010, p. 1378), whereas "Wikipedias of small and endangered languages show a high dependency on bots and a relatively small percentage of human edits. Oriya, for instance, depends 89 per cent on automated software programmes" (2010, p. 1381). They concluded that Wikipedia should be seen as "a sociotechnical system" (2010, p. 1384), and its growth should be analyzed in terms of "its human *and* non-human actors" (2010, p. 1383).

Some other micro-level elements may hinder Wikipedia's growth, such as editing conflicts, contributor retention, and local issues specific to a language version and community. Lam and Riedl (2011) investigated the growth of Wikipedia as a whole over all languages over a number of years. They observed that the number of new articles created was decreasing, and argued that the main challenge for Wikipedia would be in determining how to attract new editors and retain the existing ones, which had become more difficult due to increasing editing conflicts. Analyzing several language versions of Wikipedia as complex networks, Zlatic et al. (2006) concluded that "it is very likely that the growth process of Wikipedias is universal" (2006, p. 9), based on results from eight characteristics of article networks. Discrepancies were greatest in the Polish Wikipedia, because of the use of calendar pages, a characteristic uncommon in other language versions. Calendar pages are articles for each year including and interlinking all days and months of that year and also other years of the century.

Macro-Level Size Factors

Some studies have investigated macro-level societal factors of Wikipedia's size. Van Dijk (2009) cited the linguistic community population, literacy, Internet availability, freedom of speech, and established tradition of encyclopedias as the main factors influencing growth of Wikipedia language editions. Rask (2008) analyzed eleven Wikipedia language editions with respect to creation date, number of speakers of the language, Human Development Index, Internet users, Wikipedia contributors and edits per article; he found several correlations between these variables. For example, the Internet penetration and level of human development were correlated with the number of contributors. Other variables that may affect the size of different language editions are culture of volunteering, willingness to translate (from other language Wikipedias), and problems with non-Latin characters. Concerning the relatively small sizes of the Korean and Chinese Wikipedias, Shim and Yang (2009) suggested the competition faced by Wikipedia from other knowledge-sharing Web-services: Korean question/answering site Jisik iN and Chinese online encyclopedia Baidu Baike.

As expected, scholars have found that Wikipedia has been growing rapidly over the years. Studies of this topic have attributed many internal micro-level factors such as internal linking, freedom of restrictions for contributors, and extensive usage of bots. In explaining differences of size across various language versions, several external macro-level factors have been cited, such as the number of language speakers and rate of Internet usage among the language speakers.

Other Content Topics

Many other studies covered topics related to Wikipedia content, but do not fit well in our other categories. Some of these studies are concerned with how real entities like people, nations, and corporations are represented in Wikipedia content. Others discuss the role bots play in content development. Other topics of inquiry are also represented. These studies are presented mostly in chronological order and also in relation to each other.

Representing real entities: Investigating how large corporations are represented on Wikipedia, DiStaso and Messner (2010) used content analysis to study the corporate Wikipedia articles of ten Fortune 500 companies. They analyzed the opinion expressed, the length and the topic and how it changed over time. Among their findings was that negative topics in 2010 tended to involve legal issue, while positive topics involved corporate social responsibility or performance issues. Examining Wikipedia's articles on the histories of Singapore and the Philippines, Luyt (2011) compared their reports with the dominant historiographic records of these two nations. He found that the record of Singapore, like the dominant historiography, was fairly straightforward in reflecting the well-accepted history. In contrast, as the history of the Philippines has multiple alternative scholarly interpretations, its Wikipedia record reflected conflicting perspectives and resorted extensively to quoting historiographical sources to record different views.

Researchers are often be quite attentive to what is written on Wikipedia about themselves: In a 2011 Nature poll that recruited 840 researchers through email and social web sites, 19% responded that they check Wikipedia once or more per week for citations to themselves or their work (Reich, 2011). In comparison, 30% checked citation-counting sites (such as ISI Web of Science) and 38% checked search engines with the same frequency. The poll reported that 9% of the researchers had edited or inserted a reference to their work on Wikipedia within the last 12 months, and around 3% has edited their own Wikipedia biography. In her analysis of online philosophy resources, Elvebakk (2008) speculated that philosophers have added Wikipedia entries for themselves as a form of self-promotion.

Bots: Some studies have suggested or developed bots (software robots) to automatically extend and improve Wikipedia content. Priedhorsky et al. (2007) speculated that the extensive use of bots to find and treat vandalism has been a major reason why "the exponential increase in the probability of encountering damage was stopped" (2007, p. 268). However, they stated that it "is likely that vandals will continue working to defeat the bots, leading to an arms race" (2007, p. 268). Adar et al. (2009) described Ziggurat, a system for detecting and resolving discrepancies in Wikipedia infoboxes in different language versions of an article. Articles from different languages are aligned using cross-language links. Then, the infoboxes are aligned using a supervised learning classifier, logistic regression. Matches are detected and infoboxes gaps are filled using its match from another language version.

Other topics: Other various aspects related to Wikipedia content have also been addressed. Wagner (2005) drew from knowledge management theories to analyze Wikipedia as a knowledge creation system. By tracing the historical development of 80 articles, he found that the wiki approach to building knowledge facilitates the knowledge acquisition goals of knowledge management. Krötzsch et al. (2007) presented Semantic MediaWiki as "an extension to be integrated in Wikipedia, that allows the typing of links between articles and the specification of typed data inside the articles in an easy-to-use manner" (2007, p. 584). They argued that "Semantic Wikipedia can become a platform for technology transfer that

is beneficial both to researchers and a large number of users worldwide" (2007, p. 585). Eventually, Wikidata, a Wikimedia Foundation project launched in 2012, has emerged as a closer fulfillment of this original vision[6]. Visualising the evolution of Wikipedia articles and that of the authors contributing to them, Kimmerle et al. (2010) observed that the pattern of evolution of both articles and authors are the same, affirming co-evolution of the Wikipedia social system and the individuals' cognitive systems.

Trends in Wikipedia Content Research

In addition to our categorizations and summaries of the scholarly literature, we carefully examined various aspects of the studies that might be particularly interesting and helpful to researchers who might want to investigate Wikipedia content. To find out the major trends in Wikipedia research, we coded all 98 studies included in our review in terms of their domain of knowledge, research design and method, and the source of data they gathered and analyzed. The quantitative depictions of the analyses are provided in Table 4 to Table 16 in the appendix. (The totals listed in the "Content" row of each table are not the sum of the other rows, because some studies treat more than one sub-topic of content.) In this section, we summarize the major trends highlighted by the tables.

Our first included study on Wikipedia content was published in 2004 (Table 4). Since then, there has been rapid increase in the number of studies, until 2008, after which the number of studies per year remained around 18 to 21. (Note that our systematic review only identified studies until June 2011, as explained in the introduction.)

Overall, around 80% of research inquiries have been focused on quality issues of Wikipedia content, and the other 20% have examined size and other aspects. Within the quality research category, reliability has attracted significantly more scholarly attention than other aspects, with 30 reliability studies, compared with comprehensiveness, featured articles, and antecedents of quality with 26, 21, and 18 studies respectively. Readability and currency are among the least examined aspects with 13, and 7 studies respectively.

We identified 78 peer-reviewed journal papers and 3 doctoral theses, as well as 17 of the most significant conference papers included in our review study (Table 5). From 98 content studies, 82 studied Wikipedia as their main topic, 14 as case among others, and 2 only as sample data for a topic not specifically focused on Wikipedia itself (Table 6).

We identified the knowledge domains of the 98 studies (Table 7). There were four high-level domains: social sciences (56%), humanities (9%), natural sciences (4%) and interdisciplinary domains that crossed the first three (48%). Within the interdisciplinary area, mostly information science (26%) and computer science (13%) scholars have investigated the topic. Information scientists have examined a broad range of Wikipedia content issues, whereas computer scientists have mainly focused on reliability issues and featured articles. Within the social sciences, information systems scholars (32%) are those most interested in the topic, predominantly investigating featured articles, reliability issues, and antecedents of quality (Table 8).

In terms of theory type (Table 9), based on Gregor's (2006) classification, the research studies mostly focused on analysis and description of the Wikipedia quality aspects (79%), they sometimes proposed new designs and concrete actions (12%), and sometimes aimed for providing explanations of the phenomena investigated (12%). 61% of the research collected cross-sectional data at one point in time, while 29% examined the topic longitudinally over time; others are not empirical studies (Table 10).

---

[6] http://en.wikipedia.org/wiki/Semantic_MediaWiki#Semantic_MediaWiki_and_Wikidata

In terms of the research designs employed, from 98 studies reviewed, 44 used statistical analyses to quantitatively describe aspects of Wikipedia content, and 25 used content analysis to quantitatively analyze qualitative Wikipedia data (Table 11). One other study (Brandes, Kenis, Lerner, & van Raaij, 2009) employed network analysis and visualization to examine how collaboration between editors related to the quality of the articles. This approach proved fruitful for investigating networks of articles, how such networks evolve, and how the network position of high-quality and low quality-articles differ.

The unit of analysis in 49% of the studies is the Wikipedia article, while 25% chose to study the whole website as the unit of the investigation (Table 12). In terms of data source, 74% of the reviewed studies draw on Wikipedia pages, while 10% draw on websites like search engines, and other remaining studies employ various data sources like survey, interview, and others (Table 13). Concerning the kind of Wikipedia pages most used for research inquiries (Table 14), as can be expected, content studies have mostly drawn from Wikipedia article pages (81%), while others have drawn from other pages like article talk pages, history records, user pages, etc. Studies of featured articles have used article talk pages more frequently than other types of studies, probably because it is interesting to know how discussions on talk pages are related to the quality of such articles.

To extract data from Wikipedia (Table 15), 69% of studies have drawn from the live Wikipedia, while another 22% have employed clones of Wikipedia downloaded directly Wikipedia data dumps or provided by Wikimedia and other data sources. Notably, currency and readability studies are the only areas that have never used cloned Wikipedia data, always preferring to use the most recent version on the live website due to the nature of their inquiry. Among all language versions of Wikipedia, the English version is the most investigated (78%); this includes 59 studies that specified the English version as the focus of the study as well as 19 studies that did not explicitly specify the Wikipedia language version. However, as we only included English-language publications, this cannot be taken as an accurate reflection of all scholarly studies ever conducted. Nonetheless, English publications have reported research on many other language versions, such as Spanish, Japanese, Dutch, Italian, and others (Table 16).

The tables provided in the appendix depict all the trends we have described in more detail.

Discussion of Key Findings

The major part of the research inquiry about Wikipedia content concerns various aspects of the quality of content, including antecedents of quality, comprehensiveness, currency, readability, reliability, and featured articles. The size and growth of Wikipedia has been the other important line of inquiry. In this section, we highlight and comment on some important findings from the body of research that we have reviewed here. Since most studies included in our review have drawn on the English Wikipedia to make conclusions (see Table 16), our findings are mostly limited to the English Wikipedia, and they may not be generalizable to other language versions.

Comprehensiveness: Wikipedia content appears to continuously grow in coverage but with different paces in various domains; public interest and the interest of the contributing community play major roles here. The topics and criteria for assessing comprehensiveness are varied, and so full comparison of the results is impractical. However, such a wide variety of the fields and topics studied supports the notion of broad and highly extensive coverage of human knowledge on Wikipedia.

There is an important issue regarding the comprehensiveness research that needs to be considered by the research community. While some studies address the intrinsic completeness of each article, others focus on the more general coverage of the whole topic across the entire website. In other words, comprehensiveness sometime refers to the depth of each article and the amount of detail included, while other times it refers to the breadth of the coverage of a specific topic across the whole Wikipedia. While the former measures the amount of detail covered by articles, the latter mostly analyzes the number and

titles of Wikipedia articles covering a specific topic area. Accordingly, the studies have investigated different units of analysis and operationalizations of comprehensiveness. Differentiating between depth and breadth as the two aspects of comprehensiveness would address the issue and enhance knowledge accumulation.

Currency: This is the content quality dimension where Wikipedia is most strikingly superior to other encyclopedias, whether printed or online, because of the vast army of thousands of human contributors and bots monitoring and editing articles. However, like comprehensiveness, the currency of articles is affected by the amount of public interest to the topic. This major limitation that compromises Wikipedia's currency and coverage arises from the voluntary nature of the Wikipedia community. The community needs to adopt strategies to particularly limit such a bias towards primarily topics of interest to the public.

Readability: The readability of Wikipedia articles seems, in general, to be at least as good as its online and offline counterparts, though the writing style is not consistent across various articles. The international topics to which non-native people contribute tend to be less readable; this is primarily an issue with English, a language with a very high number of non-native speakers. The Wikipedia community could try to encourage more balance between native and non-native speakers to increase the readability of such articles, perhaps by encouraging native speakers to adopt a proof-reading and copy-editing role.

Across studies, readability is operationalized in various ways, and this might be partly responsible for the inconsistent findings. Readability is measured in terms of cognitive complexity, formality, number of passive sentences, number of words per sentence, number of sentences per paragraph, word length, sentence length, and lexical density. Developing and testing a valid and reliable formative construct including some or all of these measures would highly facilitate the future readability studies.

Reliability: Reliability or accuracy is the core to the Wikipedia content quality discussions. As we have reviewed, the studies have demonstrated mixed results, though the positive evaluations of Wikipedia are more numerous than the negative ones. However, the partial reliability of Wikipedia should not devalue the achievements of the crowd in developing Wikipedia, because any encyclopedic content is partial in nature, and it is best not to compare it with the "absolute" sources (Magnus, 2008); Wikipedia has been shown to be quite comparable with the venerable *Encyclopedia Britannica*. Moreover, Wikipedia has other epistemological properties that compensate its limitations (2008). For any encyclopedia, three epistemic aspects are essential: the amount of knowledge that can be learned, the speed by which the knowledge can be learned, and the number of people that can acquire the knowledge. Compared to other sources, Wikipedia is advantageous in the speed and number of knowledge acquirers.

One limitation to Wikipedia reliability research is the predominance of health-related studies (see Table 2), where the reliability criteria are generally stricter. More multi-disciplinary evaluations of Wikipedia reliability would give a more general picture.

Another noteworthy commentary is that talking about Wikipedia in general as being reliable or unreliable may not be very representative of what Wikipedia really is (Magnus, 2009). Rather, it would be more relevant to talk about which parts of Wikipedia are more or less reliable, and why this is the case. From this point of view, studies examining antecedents of content quality would be worthwhile. Another related issue is how to teach people to evaluate the reliability of the content they consume. This has been a growing concern for teachers and librarians. However, since this is more of a readership issue, we examine it in a separate review more focused on such issues (Okoli et al., 2014).

Antecedents of Quality: Factors affecting content quality include group characteristics like group size, diversity, participation level, etc. There are also editing patterns that enhance quality, like contribution domain, content vs. surface editing, and anonymity; public attention and public good are other features that enhance content quality.

Group characteristics are the major factors affecting the quality of content. This affirms the idea that the Wikipedia community is the primary success factor of Wikipedia. Wikipedia articles improve in quality as long as they are supported by a balanced community of new and experienced highly participative editors with shared language facilitating their cooperation. However, this should not distract Wikipedia scholars from looking into other contributing factors like the norms and practices governing group activities. Although editing patterns and processes are partially addressed, more studies of this type would be worthwhile.

Featured articles: Featured articles of Wikipedia have been examined in terms of both the specifics of the processes monitoring content quality and in terms of the factors leading to promotion or demotion of articles to or from featured status. Also, many such articles assumed that featured articles are high-quality, and used them to validate the proposed computational measures of content quality. Although it is demonstrated that the featured article evaluation processes are not perfect and may include defective articles, it has been confirmed that the quality of featured articles is significantly higher than that of others; this is supported by many studies computationally measuring the quality of content.

While most studies on this topic either validated content quality measures or looked into the factors affecting article status, there are very few that examined the vetting processes, the practices around them, and the possible deficiencies. Such studies could inform new community rules and regulations improving the vetting processes.

Size of Wikipedia: Wikipedia's size has been growing rapidly, but the rate of increase seems to be decreasing recently. Various micro- and macro-level factors have been enumerated to account for Wikipedia's huge content size. Micro-level factors include links to non-existent articles, free editing, bots, editing conflicts, and contributor retention; macro-level factors include linguistic community population, literacy, Internet availability, freedom of speech, Human Development Index, having local competing counterparts, etc.

The distinction between macro- a micro-level factors can let the community better understand Wikipedia's growth, and let them make informed decisions on how to treat the decreasing growth rate. While the macro-level factors put boundaries on Wikipedia's growth rate over which the community has little control, the micro-level factors should be the focus of continued growth strategies. Among such strategies could be promoting linking to non-existent articles, developing bots, and decreasing editing conflicts.

Conclusion

Wikipedia is a prominent source of information for many people in various fields and professions. Its popularity has instigated research inquires about various aspects of Wikipedia's content, including its quality dimensions of coverage, currency, accuracy, readability, and also its ever-growing size. While the Wikipedia phenomenon and the debates around its content is over a decade old, there is still much disagreement about different aspects of Wikipedia content. This might be due to the limited efforts to accumulate and aggregate the vast amount of scholarly discoveries on Wikipedia content in various fields from health, history, and philosophy, to information and engineering sciences.

In addressing this issue, this present paper has contributed to the body of knowledge in two main ways. First, it has reported on part of a systematic review of scholarly research on Wikipedia that identified and presented the streams of research on various aspects of Wikipedia content; such aspects include quality dimensions and the size of Wikipedia. Analyzing content quality dimensions not only informs the current state of knowledge, but also highlights the issues that remain open or unattended for future studies. Also, such a presentation of the current understanding of Wikipedia content may inform the future strategies and policies of the Wikipedia community. As a secondary contribution, this paper has analyzed the trends

in various aspects of Wikipedia content studies such as the number of studies, the fields of inquiry, the research methods employed, data gathering and analysis methods used. Such analysis informs future scholars of the research methods employed, and facilitates designing effective inquiries of Wikipedia content.

Overall, Wikipedia content research is very broad, extensively covering various aspects of Wikipedia. However, such diversity has made it difficult for the scholarly community to locate the relevant research and to build new research based on and consistent with previous studies. We believe that this systematic review addresses this issue and significantly helps researchers to accumulate knowledge in this important area.

Acknowledgments

References

Adamic, L. A., Xiao Wei, Jiang Yang, Gerrish, S., Nam, K. K., & Clarkson, G. S. (2010). Individual focus and knowledge contribution. *First Monday*, *15*(3). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/2841/2475

Adar, E., Skinner, M., & Weld, D. S. (2009). Information arbitrage across multi-lingual Wikipedia. In *2nd ACM International Conference on Web Search and Data Mining, WSDM'09, February 9, 2009 - February 12, 2009* (pp. 94–103). Barcelona, Spain: Association for Computing Machinery. Retrieved from http://dx.doi.org/10.1145/1498759.1498813

Altmann, U. (2005). Representation of Medical Informatics in the Wikipedia and its Perspectives. *Studies in Health Technology and Informatics*, *116*, 755–760.

Anthony, D., Smith, S. W., & Williamson, T. (2009). Reputation and Reliability in Collective Goods. *Rationality and Society*, *21*(3), 283–306. doi:10.1177/1043463109336804

Arazy, O., & Kopak, R. (2011). On the measurability of information quality. *Journal of the American Society for Information Science and Technology*, *62*(1), 89–99. doi:10.1002/asi.21447

Arazy, O., Nov, O., Patterson, R., & Yeo, L. (2011). Information Quality in Wikipedia: The Effects of Group Composition and Task Conflict. *Journal of Management Information Systems*, *27*(4), 71–98.

Bekker-Nielsen, T. (2011). Historie på Wikipedia. *Noter*, *188*, 48–52.

Blumenstock, J. E. (2008). Size Matters: Word Count as a Measure of Quality on Wikipedia. In *Proceedings of the 17th International World Wide Web Conference (WWW2008). April 21-25, 2008. Beijing, China*.

Bragues, G. (2009). Wiki-Philosophizing in a Marketplace of Ideas: Evaluating Wikipedia's Entries on Seven Great Minds. *MediaTropes eJournal*, *2*(1), 117–158.

Brandes, U., Kenis, P., Lerner, J., & van Raaij, D. (2009). Network analysis of collaboration structure in Wikipedia. In *Proceedings of the 18th international conference on World wide web* (pp. 731–740). New York, NY, USA: ACM. doi:10.1145/1526709.1526808

Brown, A. R. (2011). Wikipedia as a Data Source for Political Scientists: Accuracy and Completeness of Coverage. *PS: Political Science & Politics*, *44*(02), 339–343. doi:10.1017/S1049096511000199

Caddick, S. (2006). Wiki and other ways to share learning online. *Nature*, *442*(7104), 744–744. doi:10.1038/442744c

Carillo, K., & Okoli, C. (2011). Generating quality open content: A functional group perspective based on the time, interaction, and performance theory. *Information & Management*, *48*(6), 208–219. doi:doi: 10.1016/j.im.2011.04.004

Chesney, T. (2006). An empirical examination of Wikipedia's credibility. *First Monday*, *11*(11). Retrieved from http://firstmonday.org/article/view/1413/1331

CityNews.ca. (2008, November 26). Why Wikipedia Is Wrong When It Comes To Prescription Medicine. *CityNews*. Toronto. Retrieved from http://www.citynews.ca/2008/11/26/why-wikipedia-is-wrong-when-it-comes-to-prescription-medicine/

Clark, M., Ruthven, I., & Holt, P. O. (2009). The evolution of genre in Wikipedia. *Journal for Language Technology and Computational Linguistics*, *24*(1), 1–22.

Clauson, K. A., Polen, H. H., Boulos, M. N. K., & Dzenowagis, J. H. (2008). Scope, completeness, and accuracy of drug information in Wikipedia. *The Annals of Pharmacotherapy*, *42*(12), 1814–1821. doi:10.1345/aph.1L474

Dalby, A. (2007). Wikipedia(s) on the language map of the world. *English Today*, *23*(02), 3. doi:10.1017/S0266078407002027

Den Besten, M., & Dalle, J.-M. (2008). Keep it simple: A companion for Simple Wikipedia? *Industry and Innovation*, *15*(2), 169–178.

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

DiStaso, M. W., & Messner, M. (2010). Forced transparency: Corporate image on Wikipedia and what it means for public relations. *Public Relations Journal*, *4*(2), 1–23.

Dondio, P., & Barrett, S. (2007). Computational trust in Web content quality: a comparative evalutation on the Wikipedia project. *Informatica*, *31*(2), 151–60.

Duguid, P. (2006). Limits of self-organization: Peer production and "laws of quality. *First Monday*, *11*(10). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/1405/1323

Ehmann, K., Large, A., & Beheshti, J. (2008). Collaboration in context: comparing article evolution among subject disciplines in Wikipedia. *First Monday*, *13*(10). Retrieved from http://firstmonday.org/article/view/2217/2034

Elia, A. (2009). Quantitative data and graphics on lexical specificity and index of readability: The case of Wikipedia. *RaeL: Revista Electronica de Linguistica Aplicada*, (8), 248–271. doi:Article

Elvebakk, B. (2008). Philosophy democratized? A comparison between Wikipedia and two other Web-based philosophy resources. *First Monday*, *13*(2). Retrieved from http://firstmonday.org/article/view/2091/1938

Emigh, W., & Herring, S. C. (2005). Collaborative Authoring on the Web: A Genre Analysis of Online Encyclopedias. In *Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) - Track 4 - Volume 04* (p. 99.1). Washington, DC, USA: IEEE Computer Society. doi:10.1109/HICSS.2005.149

Fallis, D. (2008). Toward an Epistemology of Wikipedia. *Journal of the American Society for Information Science and Technology*, *59*(10), 1662–1674. doi:10.1002/asi.20870

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, *438*(7070), 900–901. doi:10.1038/438900a

Goldspink, C. (2009). Social self-regulation in computer mediated communities: the case of Wikipedia. *International Journal of Agent Technologies & Systems*, *1*(1), 19–33.

Goldspink, C. (2010). Normative behaviour in Wikipedia. *Information*, *13*(5), 652–673.

Greenstein, S. (2007). Wagging Wikipedia's long tail. *IEEE Micro*, *27*(2), 6+79.

Gregor, S. (2006). The nature of theory in information systems. *MIS Q.*, *30*(3), 611–642.

Haigh, C. A. (2010). Wikipedia as an evidence source for nursing and healthcare students. *Nurse Education Today*. doi:10.1016/j.nedt.2010.05.004

Halavais, A., & Lackaff, D. (2008). An analysis of topical coverage of Wikipedia. *Journal of Computer Mediated Communication*, *13*(2), 429–440. doi:10.1111/j.1083-6101.2008.00403.x

Harding, A. (2008, November 25). Wikipedia often omits important drug information: study. *Reuters*. New York. Retrieved from http://www.reuters.com/article/2008/11/25/us-wikipedia-information-idUSTRE4AN7BO20081125

Huvila, I. (2010). Where does the information come from? Information source use patterns in Wikipedia. *Information Research*, *15*(3), 24 pp.

Jancarik, A., & Jancarikova, K. (2010). Wiki Tools in the Preparation and Support of e-Learning Courses. *Electronic Journal of E-Learning*, *8*(2), 123–32.

Jones, J. (2008). Patterns of revision in online writing: A study of Wikipedia's featured articles. *Written Communication*, *25*(2), 262–289. doi:10.1177/0741088307312940

Jullien, N. (2012). What we know about Wikipedia. A review of the literature analyzing the project(s). *Available at SSRN 2053597*. Retrieved from https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID2053597_code728676.pdf?abstractid=2053597&mirid=2

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

Kimmerle, J., Moskaliuk, J., Harrer, A., & Cress, U. (2010). VISUALIZING CO-EVOLUTION OF INDIVIDUAL AND COLLECTIVE KNOWLEDGE. *Information, Communication & Society*. Retrieved from http://www.informaworld.com/10.1080/13691180903521547

Kittur, A., Chi, E. H., & Suh, B. (2009). What's in Wikipedia?: mapping topics and conflict using socially annotated category structure. In *Proceedings of the 27th international conference on Human factors in computing systems* (pp. 1509–1512). New York, NY, USA: ACM. doi:10.1145/1518701.1518930

Kittur, A., & Kraut, R. E. (2008). Harnessing the wisdom of crowds in wikipedia: Quality through coordination. In *2008 ACM Conference on Computer Supported Cooperative Work, CSCW 08, November 8, 2008 - November 12, 2008* (pp. 37–46). San Diego, CA, United states: Association for Computing Machinery. doi:10.1145/1460563.1460572

Klemp, N. J., & Forcehimes, A. T. (2010). From Town-Halls to Wikis: Exploring Wikipedia's Implications for Deliberative Democracy. *Journal of Public Deliberation*, *6*(2), 4.

Korosec, L., Limacher, P. A., Lüthi, H. P., & Brändle, M. P. (2010). Chemical Information Media in the Chemistry Lecture Hall: A Comparative Assessment of Two Online Encyclopedias. *CHIMIA International Journal for Chemistry*, *64*(5), 309–314. doi:10.2533/chimia.2010.309

Krötzsch, M., Vrandecic, D., Volkel, M., Haller, H., & Studer, R. (2007). Semantic Wikipedia. *Web Semantics*, *5*(4), 251–261. doi:10.1016/j.websem.2007.09.001

Lam, S. T. K., & Riedl, J. (2011). The past, present, and future of Wikipedia. *Computer*, *44*(3), 87–90.

Lam, S. T. K., Uduwage, A., Dong, Z., Sen, S., Musicant, D. R., Terveen, L., & Riedl, J. (2011). WP:Clubhouse? An exploration of Wikipedia's gender imbalance. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (pp. 1–10). New York, NY, USA: ACM. doi:10.1145/2038558.2038560

Lanamäki, A., Okoli, C., Mehdi, M., & Mesgari, M. (2011). Protocol for Systematic Mapping of Wikipedia Studies. In *Proceedings of IRIS 2011 – The 34th Information Systems Research Seminar in Scandinavia*. Turku, Finland.

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. doi:10.1197/jamia.M3059

Lavsa, S. M., Corman, S. L., Culley, C. M., & Pummer, T. L. (2011). Reliability of Wikipedia as a medication information source for pharmacy students. *Currents in Pharmacy Teaching and Learning*, *3*(2), 154–158. doi:10.1016/j.cptl.2011.01.007

Leithner, A., Maurer-Ertl, W., Glehr, M., Friesenbichler, J., Leithner, K., & Windhager, R. (2010). Wikipedia and osteosarcoma: a trustworthy patients' information? *Journal of the American Medical Informatics Association*, *17*(4), 373–374. doi:10.1136/jamia.2010.004507

Lewandowski, D., & Spree, U. (2011). Ranking of Wikipedia articles in search engines revisited: Fair ranking for reasonable quality? *Journal of the American Society for Information Science and Technology*, *62*(1), 117–132. doi:10.1002/asi.21423

Lih, A. (2004). Wikipedia as participatory Journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource. In *5th International Symposium on Online Journalism*. Retrieved from http://jmsc.hku.hk/faculty/alih/publications/utaustin-2004-wikipedia-rc2.pdf

Lindsey, D. (2010). Evaluating quality control of Wikipedia's feature articIes. *First Monday*, *15*(4). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/2721/2482

Luyt, B. (2011). The nature of historical representation on Wikipedia: Dominant or alterative historiography? *Journal of the American Society for Information Science and Technology*, *62*(6), 1058–1065. doi:10.1002/asi.21531

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 (Print)*, *61*(4), 715–722. doi:10.1002/asi.21304

Luyt, B., Tay, C. H. A., Lim, H. T., & Cheng, K. H. (2008). Improving wikipedia's accuracy: Is edit age a solution? *Journal of the American Society for Information Science and Technology*, *59*(2), 318–330. doi:10.1002/asi.20755

Magnus, P. D. (2008). Early response to false claims in Wikipedia. *First Monday*, *13*(9). Retrieved from http://firstmonday.org/article/view/2115/2027

Magnus, P. D. (2009). On Trusting Wikipedia. *Episteme - Edinburgh*, *6*(1).

Martin, O. S. (2010). A Wikipedia Literature Review. *arXiv:1110.5863*. Retrieved from http://arxiv.org/abs/1110.5863

McGuinness, D. L., Zeng, H., Silva, P. P. da, Ding, L., Narayanan, D., & Bhaowal, M. (2006). Investigations into Trust for Collaborative Information Repositories: A Wikipedia Case Study. In *Proceedings of the Workshop on Models of Trust for the Web*. Edinburgh, United Kingdom. Retrieved from http://ebiquity.umbc.edu/_file_directory_/papers/274.pdf

Mercer, J. (2007). Wikipedia and "open source" mental health information. *Scientific Review of Mental Health Practice*, *5*(1), 88–92. doi:Article

Milne, D., Medelyan, O., & Witten, I. H. (2006). Mining Domain-Specific Thesauri from Wikipedia: A Case Study. In *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 442–448). doi:10.1109/WI.2006.119

Monaci, S. (2009). Quality assessment process in Wikipedia's Vetrina: the role of the community's policies and rules. *Observatorio (OBS*)*, *3*(1). Retrieved from http://obs.obercom.pt/index.php/obs/article/viewArticle/240

Niederer, S., & Dijck, J. van. (2010). Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. *New Media & Society*, *12*(8), 1368–1387. doi:10.1177/1461444810365297

Nielsen, F. Å. (2007). Scientific citations in Wikipedia. *First Monday*, *12*(8). Retrieved from http://firstmonday.org/article/view/1997/1872

Nielsen, F. Å. (2008). Clustering of scientific citations in Wikipedia. In *Wikimania*. Retrieved from http://arxiv.org/abs/0805.1154

Nielsen, F. Å. (2012, February 6). Wikipedia Research and Tools: Review and Comments. SSRN. Retrieved from http://ssrn.com/abstract=2129874

Okoli, C. (2009). A brief review of studies of Wikipedia in peer-reviewed journals. In *Digital Society, 2009. ICDS'09. Third International Conference on* (pp. 155–160). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4782868

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 American Society for Information Science and Technology*, *Forthcoming*.

Okoli, C., Mehdi, M., Mesgari, M., Nielsen, F., & Lanamäki, A. (2012, October 24). *The People's Encyclopedia Under the Gaze of the Sages: A Systematic Review of Scholarly Research on Wikipedia*. SSRN Scholarly Paper, Montreal. Retrieved from http://papers.ssrn.com/abstract=2021326

Okoli, C., & Schabram, K. (2009a). Protocol for a systematic literature review of research on the Wikipedia. In *Proceedings of the International Conference on Management of Emergent Digital EcoSystems (MEDES)* (p. 73). Lyon, France: Association for Computing Machinery. Retrieved from http://dl.acm.org/citation.cfm?id=1643912

Okoli, C., & Schabram, K. (2009b). Protocol for a systematic literature review of research on the Wikipedia. *Sprouts: Working Papers in Information Systems*, *9*(65). Retrieved from http://sprouts.aisnet.org/9-65

Okoli, C., & Schabram, K. (2010). A guide to conducting a systematic literature review of information systems research. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1954824

Okoli, C., Schabram, K., & Kader, B. A. (2009). From the Academy to the Wiki: Practical Applications of Scholarly Research on Wikipedia. In *Proceedings of Wikimania*. Buenos Aires: Wik. Retrieved from http://chitu.okoli.org/images/stories/bios/pro/research/open/Okolietal2009Wikimania.pdf

Ortega, F. (2009). *Wikipedia. A quantitative analysis* (Doctoral thesis). Universidad Rey Juan Carlos, Madrid, Spain. Retrieved from http://libresoft.es/Members/jfelipe/thesis-wkp-quantanalysis/view

Page, R. (2010). Wikipedia as an encyclopaedia of life. *Organisms Diversity & Evolution*, *10*(4), 343–349. doi:10.1007/s13127-010-0028-9

Pender, M. P., Lasserre, K. E., Kruesi, L. M., Mar, C. D., & Anuradha, S. (2008). Putting Wikipedia to the test: a case study. In *The Special Libraries Association Annual Conference*. Retrieved from http://espace.library.uq.edu.au/eserv/UQ:193433/SLA_Paper.pdf

Poderi, G. (2009). Comparing featured article groups and revision patterns correlations in Wikipedia. *First Monday*, *14*(5). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/2365/2182

Priedhorsky, R., Chen, J., Lam, S. T. K., Panciera, K., Terveen, L., & Riedl, J. (2007). Creating, destroying, and restoring value in wikipedia. In *2007 International ACM Conference on Supporting Group Work, GROUP'07, November 4, 2007 - November 7, 2007* (pp. 259–268). Sanibel Island, FL, United states: Association for Computing Machinery. doi:10.1145/1316624.1316663

Purdy, J. P. (2009). When the Tenets of Composition Go Public: A Study of Writing in Wikipedia. *College Composition and Communication*, *61*.

Radtke, P. J., & Munsell, J. F. (2010). Wikipedia as a tool for forestry outreach. *Journal of Forestry*, *108*(7), 354–359.

Rahman, M. M. (2006). *Essays analyzing blogs and Wikipedia*. The University of Kansas, United States - - Kansas. Retrieved from http://proquest.umi.com/pqdweb?did=1126778281&Fmt=7&clientId=10306&RQT=309&VName=PQD

Rahman, M. M. (2008). An Analysis of Wikipedia. *Journal of Information Technology Theory and Application (JITTA)*, *9*(3), 81.

Rajagopalan, M. S., Khanna, V. K., Leiter, Y., Stott, M., Showalter, T. N., Dicker, A. P., & Lawrence, Y. R. (2011). Patient-oriented cancer information on the Internet: a comparison of Wikipedia and a

professionally maintained database. *Journal of Oncology Practice*, *7*(5), 319–323. doi:10.1200/JOP.2010.000209

Rajagopalan, M. S., Khanna, V., Stott, M., Leiter, Y., Showalter, T., Dicker, A., & Lawrence, Y. (2010). Accuracy of cancer information on the Internet: A comparison of a Wiki with a professionally maintained database. *Bodine Journal*, *3*(1), 7s.

Ransbotham, S., & Kane, G. C. (2011). Membership Turnover and Collaboration Success in Online Communities: Explaining Rises and Falls from Grace in Wikipedia. *MIS Quarterly*, *35*(3), 613–627.

Rask, M. (2008). The reach and richness of Wikipedia: Is Wikinomics only for rich countries. *First Monday*, *13*(6). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/2046/1970

Reagle, J., & Rhue, L. (2011). Gender Bias in Wikipedia and Britannica. *International Journal of Communication*, *5*(0), 21.

Rector, L. H. (2008). Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles. *Reference Services Review*, *36*(1), 7–22.

Reich, E. S. (2011). Online reputations: Best face forward. *Nature News*, *473*(7346), 138–139. doi:10.1038/473138a

Rosenzweig, R. (2006). Can History Be Open Source? Wikipedia and the Future of the Past. *Journal of American History*, *93*(1), 117–146. doi:Article

Royal, C., & Kapila, D. (2009). What's on Wikipedia, and what's not … ? Assessing completeness of information. *Social Science Computer Review*, *27*(1), 138–148. doi:10.1177/0894439308321890

Rush, E. K., & Tracy, S. J. (2010). Wikipedia as Public Scholarship: Communicating Our Impact Online. *Journal of Applied Communication Research*, *38*(3), 309–315. doi:10.1080/00909882.2010.490846

Santana, A., & Wood, D. J. (2009). Transparency and social responsibility issues for Wikipedia. *Ethics and Information Technology*, *11*(2), 133–144. doi:10.1007/s10676-009-9193-y

Schweitzer, N. J. (2008). Wikipedia and Psychology: Coverage of Concepts and Its Use by Undergraduate Students. *Teaching of Psychology*, *35*(2), 81–85. doi:10.1080/00986280802004594

Shim, J. P., & Yang, J. (2009). Why is Wikipedia not more widely accepted in Korea and China? factors affecting knowledge-sharing adoption. *Decision Line*, *40*(2), 12–15.

Slashdot. (2004, July 28). Wikipedia Founder Jimmy Wales Responds. *Slashdot*. Retrieved from http://slashdot.org/story/04/07/28/1351230/wikipedia-founder-jimmy-wales-responds

Smith, D. M. D., Onnela, J.-P., & Johnson, N. F. (2007). Accelerating networks. *New Journal of Physics*, *9*(181). doi:10.1088/1367-2630/9/6/181

Spinellis, D., & Louridas, P. (2008). The collaborative organization of knowledge. *Communications of the ACM*, *51*(8), 68–73.

Stankus, T., & Spiegel, S. E. (2010a). 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. doi::10.1080/01942620903579435

Stankus, T., & Spiegel, S. E. (2010b). Wikipedia, Scholarpedia, and References to Journals in the Brain and Behavioral Sciences: A Comparison of Cited Sources and Recommended Readings in

Matching Free Online Encyclopedia Entries. *Science & Technology Libraries*, *29*(3), 258–265. doi:10.1080/0194262X.2010.497711

Stvilia, B., Al-faraj Abdullah, & Yi, Y. J. (2009). Issues of cross-contextual information quality evaluation—The case of Arabic, English, and Korean Wikipedias. *Library & Information Science Research*, *31*(4), 232–239. doi:10.1016/j.lisr.2009.07.005

Stvilia, B., & Gasser, L. (2008). An activity theoretic model for information quality change. *First Monday*, *13*(4). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/2126/1951

Stvilia, B., Gasser, L., Twidale, M. B., & Smith, L. C. (2007). A framework for information quality assessment. *Journal of the American Society for Information Science and Technology*, *58*(12), 1720–1733. doi:10.1002/asi.v58:12

Stvilia, B., Twidale, M. B., Smith, L. C., & Gasser, L. (2008). Information quality work organization in Wikipedia. *Journal of the American Society for Information Science and Technology*, *59*(6), 983–1001. doi:10.1002/asi.20813

Van Dijk, Z. (2009). Wikipedia and lesser-resourced languages. *Language Problems & Language Planning*, *33*(3), 234–250. doi:10.1075/lplp.33.3.03van

Vassia Atanassova. (2011). Topics of Bioengineering in Wikipedia. *Bioautomation*, *13*. Retrieved from http://www.biomed.bas.bg/bioautomation/2009/vol_13.3/files/13.3_4.1.pdf

Viégas, F. B., Wattenberg, M., & McKeon, M. M. (2007). The Hidden Order of Wikipedia. In *Lecture Notes in Computer Science* (Vol. 4564, pp. 445–454). Berlin/Heidelberg: Springer. doi:10.1007/978-3-540-73257-0_49

Voss, J. (2005). Measuring Wikipedia. In *Proceedings International Conference of the International Society for Scientometrics and Informetrics : 10th*. Retrieved from http://eprints.rclis.org/archive/00003610/

Wagner, C. (2005). Breaking the knowledge acquisition bottleneck through conversational knowledge management. *Information Resources Management Journal*, *19*(1), 70–83.

Wedemeyer, B., Yakubova, N., Kallenbach, J., Ekdahl, A., Lesko, L., Reed, E., & Schwartz, K. (2008). Quality of the science articles on the English Wikipedia: Preliminary results. In *Wikimania 2008*. Alexandria, Egypt. Retrieved from http://www.youtube.com/watch?v=B7bCZbHHeZI

West, K., & Williamson, J. (2009). Wikipedia: friend or foe? *Reference Services Review*, *37*(3), 260–271.

Wilkinson, D. M., & Huberman, B. A. (2007). Assessing the value of cooperation in Wikipedia. *First Monday*, *12*(4). Retrieved from http://firstmonday.org/article/view/1763/1643

Willinsky, J. (2008). Socrates Back on the Street: Wikipedia's Citing of the "Stanford Encyclopedia of Philosophy", *2*, 1269–88.

Zeng, H., Alhossaini, M. A., Ding, L., Fikes, R., & McGuinness, D. L. (2006). Computing trust from revision history. In *STAR. Vol. 44* (Vol. 44). Retrieved from http://ebiquity.umbc.edu/_file_directory_/papers/302.pdf

Zlatic, V., Bozicevic, M., tefancic, H., & Domazet, M. (2006). Wikipedias: Collaborative web-based encyclopedias as complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *74*(1), 016115. doi:10.1103/PhysRevE.74.016115

Appendix: Descriptive Statistics of Studies Examining Wikipedia Content

Table 4. The number of Wikipedia content related studies over the years

	2004	2005	2006	2007	2008	2009	2010	2011	2012	Total Sum of Studies
Content	**1**	**4**	**8**	**13**	**21**	**18**	**19**	**12**	**2**	**98**
Quality	**1**	**2**	**7**	**9**	**19**	**16**	**16**	**11**	**1**	**82**
Antecedents of quality	1		3	1	4	2	2	5		**18**
Comprehensiveness			2	1	7	5	6	5		**26**
Currency				1	3	2	1			**7**
Featured articles			2	5	3	6	2	3		**21**
Readability and style		1		2	3	5	1	1		**13**
Reliability		1	2	2	9	3	8	4	1	**30**
Size of Wikipedia	**1**	**1**	**2**	**3**	**2**	**2**	**1**	**1**		**13**
Other content topics		1		2	1	1	2	1	1	**9**

Table 5. The publication type of Wikipedia content related studies

	Conference paper	Journal article	Thesis
Content	**17**	**78**	**3**
Quality	**13**	**66**	**3**
Antecedents of quality	3	14	1
Comprehensiveness	3	23	
Currency	1	6	
Featured articles	6	14	1
Readability and style	2	11	
Reliability	2	26	2
Size of Wikipedia	**2**	**9**	**2**
Other content topics	**3**	**6**	

Table 6. The Wikipedia coverage of Wikipedia content related studies

	Case	Main topic	Sample data
Content	**14**	**82**	**2**
Quality	**11**	**69**	**2**
Antecedents of quality	1	16	1
Comprehensiveness	5	21	
Currency	1	6	
Featured articles	2	18	1
Readability and style	5	8	
Reliability	4	26	
Size of Wikipedia	**1**	**12**	
Other content topics	**2**	**7**	

Table 7. The domains of Wikipedia content related studies

	Humanities	Interdisciplinary	Natural sciences	Social sciences
Content	**9**	**48**	**4**	**56**
Quality	**8**	**40**	**3**	**47**
Antecedents of quality	2	4		12
Comprehensiveness	2	17	2	8
Currency	1	2		4
Featured articles		9		12
Readability and style	2	6		5
Reliability	3	18	1	12
Size of Wikipedia	**1**	**5**	**1**	**7**
Other content topics	**2**	**4**		**5**

Table 8. The detailed domains of Wikipedia content related studies

	Biology	Business	Chemistry	Communications	Computer science	Economics	Education	Health	History	Information science	Information systems	Journalism	Knowledge management	Library science	Linguistics	Philosophy and ethics	Physics	Political science	Rhetoric	Sociology
Content	**2**	**2**	**2**	**2**	**13**	**1**	**6**	**10**	**2**	**26**	**32**	**1**	**3**	**7**	**2**	**4**	**2**	**2**	**1**	**3**
Quality	**2**	**1**	**2**	**2**	**9**		**6**	**10**	**2**	**21**	**28**		**1**	**7**	**1**	**4**	**1**	**2**	**1**	**1**
Antecedents of quality		1		1	1				1	3	8			1		1		1		
Comprehensiveness	1		1	1	1		2	6	1	10	2			3		1				
Currency							2	1		1	1			1		1				
Featured articles					4					5	11									1
Readability and style							2	2		4	3				1				1	
Reliability	1		1		4			8	2	6	8		1	3		1	1	1		
Size of Wikipedia					1	1				**4**	**3**	**1**	**1**		**1**		**1**			**1**
Other content topics		1			4				1	1	2		1			1				1

Table 9. The theory type of Wikipedia content related studies

	Analysis	Design and action	Explanation
Content	**79**	**12**	**12**
Quality	**68**	**9**	**10**
Antecedents of quality	13		5
Comprehensiveness	23	2	2
Currency	5	1	1
Featured articles	11	6	4
Readability and style	12	1	
Reliability	28	2	1
Size of Wikipedia	**8**	**1**	**4**
Other content topics	**6**	**3**	

Table 10. The time dimension of Wikipedia content related studies

	Cross-sectional	Longitudinal	N/A
Content	**61**	**29**	**8**
Quality	**54**	**23**	**5**
Antecedents of quality	4	10	4
Comprehensiveness	24	2	
Currency	6	1	
Featured articles	9	12	
Readability and style	10	3	
Reliability	24	5	1
Size of Wikipedia	**3**	**5**	**5**
Other content topics	**6**	**3**	

Table 11. The research design of Wikipedia content related studies

	Action research	Case study	Conceptual	Content analysis	Design science	Discourse analysis	Econometrics and time series	Experiment	Historical analysis	Mathematical modeling	Phenomenology	Statistical analysis
Content	1	6	8	25	1	1	2	8	1	8	2	44
Quality	1	4	5	23		1	2	6	1	7	2	36
Antecedents of quality		1	2	3			2	2		3	1	4
Comprehensiveness		2	2	8							1	13
Currency				2								5
Featured articles		3		6		1		1		3		8
Readability and style		1		6				2			1	3
Reliability	1	1	1	9				3	1	1	1	14
Size of Wikipedia		**1**	**3**							**2**		**7**
Other content topics		**1**		**2**	**1**			**2**	**1**	**1**		**3**

Table 12. The unit of analysis of Wikipedia content related studies

	Article	Article view	Category	Edit	Language	Subject	User	Website	N/A
Content	**49**	**1**	**2**	**10**	**5**	**6**	**9**	**25**	**7**
Quality	**43**		**2**	**9**	**2**	**6**	**7**	**18**	**6**
Antecedents of quality	11			2		1		1	3
Comprehensiveness	8		1			4	2	10	1
Currency	3			1				3	
Featured articles	17			3			4		
Readability and style	6		1	2	1			2	1
Reliability	14			1	1	3	1	9	3
Size of Wikipedia	**2**			**1**	**2**		**1**	**6**	**2**
Other content topics	**6**	**1**			**1**		**1**	**2**	

Table 13. The data source of Wikipedia content related studies

	Archival records	Computer usage logs	Direct observation	Documents	Experiment responses	Interview responses	Survey responses	Websites	Wikipedia pages	N/A
Content	**2**	**2**	**1**	**2**	**9**	**3**	**5**	**10**	**73**	**8**
Quality	**2**		**1**	**2**	**6**	**3**	**5**	**7**	**63**	**5**
Antecedents of quality					2	1		1	10	4
Comprehensiveness	1			1			2	4	19	
Currency								1	6	
Featured articles	1		1		1	1	1		17	
Readability and style					1		1	3	9	
Reliability				1	3	1	2	4	20	1
Size of Wikipedia					**1**			**2**	**7**	**5**
Other content topics		**2**			**2**			**1**	**5**	

Table 14. The Wikipedia page type used for data extraction across Wikipedia content related studies

	Article	Article:talk	Discussion and Q&A	History	Information categorization and navigation	Other	Policy	Quality management	User	User:talk	N/A
Content	**81**	**5**	**2**	**4**	**2**	**3**	**1**	**2**	**1**	**2**	**12**
Quality	**69**	**5**	**2**	**2**	**1**	**3**	**1**	**2**	**1**	**2**	**8**
Antecedents of quality	14						1			2	4
Comprehensiveness	23					2			1		1
Currency	5			1							1
Featured articles	14	4	1	1				2			
Readability and style	11					1					1
Reliability	26	1	1		1						2
Size of Wikipedia	**8**			**1**	**1**						**5**
Other content topics	**8**			**1**							**1**

Table 15. The data extraction type of Wikipedia content related studies

	Clone	Live Wikipedia	N/A
Content	**22**	**69**	**7**
Quality	**16**	**61**	**5**
Antecedents of quality	5	10	3
Comprehensiveness	2	24	
Currency		7	
Featured articles	9	12	
Readability and style		13	
Reliability	4	24	2
Size of Wikipedia	**4**	**6**	**3**
Other content topics	**3**	**6**	

Table 16. The Wikipedia language versions studies across Wikipedia content related studies

	All languages	Arabic	Croatian	Czech	Dutch	English	German	Italian	Japanese	Not specified	Simple English	Spanish	Multiple	N/A
Content	1	1	1	1	1	59	5	1	2	19	1	1	7	4
Quality		1		1	1	53	1	1		18	1		2	3
Antecedents of quality					1	11				5				1
Comprehensiveness				1		22	1			2				
Currency						7								
Featured articles		1				11		1		7			1	
Readability and style						9	1			1	1		1	
Reliability						20				7			1	2
Size of Wikipedia	1		1			3	2		2	2			5	1
Other content topics						6	2			1		1	1	