"Timely Quality Problem Resolution in Peer-Production Systems:
The Impact of Bots, Policy Citations, and Contributor Experience"

by

Vitali Mindel, Aleksi Aaltonen, Arun Rai, Lars Mathiassen, and Wael Jabr

published in *Information Systems Research*

Online Appendixes

EC1 Literature on Wikipedia Content Quality

We reviewed literature on content quality in Wikipedia to develop a focused research model and to justify the theoretical and practical importance of our dependent variable.

Table EC-1. Wikipedia Content Quality Studies

Study	Theoretical perspective	Method and data	Unit of analysis	General findings	Dependent variable	Specific findings
Anderka et al. (2012)	An inductive machine-learning-based prediction model of quality problems; no specific hypotheses	Machine-learning density estimation with class probability estimation	10,000 Wikipedia articles with one or more out of 10 types of cleanup templates	A method for mining and extracting cleanup templates; the algorithm successfully predicted quality issues associated with 4 out of 10 templates	Cleanup tag (template) as an indication of quality problems	N/A
Anthony et al. (2009)	Deductive hypothesis development partially based on previous empirical findings	The examination of the effect of contributor anonymity and edit history on the quality of contributions in Wikipedia	The quality of contributions measured as the rate of each contributor's content retained in the current version	Registered users' quality increases with more contributions; a large number of high-quality contributions come from anonymous contributors who contribute only once	The retention rate of contributions measured as the proportion of characters retained per contribution by each contributor	*Production participants*: Calls into question the importance of equitability and veterans *Production organization*: Veteran registered contributors improve the quality of contributions with experience although one-off unregistered contributors provide the most single high-quality edits. Implication to our paper: work concentration in one-off contributions produce highest quality.
Arazy and Kopak (2011)	Information quality as accuracy, completeness, objectivity, and representation	270 students assessing the quality of 2 Wikipedia articles out of 100 along four dimensions and a composite information quality	The information quality of a Wikipedia article along four dimensions	Some dimensions (e.g., accuracy) are easier to agree by different assessors than others (e.g. completeness); the dimensions lend themselves differently to measurement; quality is difficult to measure reliably	Accuracy, completeness, objectivity, representation, and composite information quality on a Likert scale; interrater agreement using intraclass correlation and the reliability of scale metrics	N/A
Arazy and Nov (2010)	Deductive hypothesis development partially based on previous empirical findings	A qualitative assessment of the accuracy and completeness of 50 articles by experts; quantitative measures of editing activities on those articles	Wikipedia article and editors	Coordination among contributors positively impacts article quality; unequal contribution effort among editors on the article level is not associated with quality; unequal contribution among the same	Article quality as assessed by experts who rated a sample of articles on a Likert scale of 1 to 5 for accuracy and completeness	*Production organization*: High GINI is positively associated with word count on a talk page which is a proxy for coordination. Word count on the talk page is associated with quality but the GINI score is not directly affecting quality. Finds that low GINI is associated with faster quality resolution and do

				editors across Wikipedia is positively associated with quality; the number of editors is not directly associated with quality		not examine how it affects word count on the talk page during a quality problem event.
Arazy et al. (2011)	Develops a limited scope theoretical model based on previous research and a theory of group work to examine how group composition, specifically editor's orientation (content providers or admins), cognitive diversity (the extent to which contributors are active on other articles as well), and task conflict (as seen in discussions on the talk page), impact quality	Randomly sampled 15–17 articles from 6 different categories; mixed methods: both quantitative and qualitative analysis	96 Wikipedia articles	The presence of admins helps resolve conflicts; a higher proportion of content-oriented contributors with diverse interests improves quality	Use Lee et al. (2002) information quality categories: accuracy and objectivity, completeness, and representation. Used three librarians to assign quality scores after the manual evaluation of the articles	*Production participants*: Differentiates between administration and content-oriented participants. Those who have an administrative title and contribute across Wikipedia are administration oriented. Those who mostly focus on the article, are content oriented. Also make a distinction between participants who participate in many articles and those who participate in few. Those who participate in many, are considered more cognitively diverse. This is another example of how research differentiates between type of participants.
Arazy et al. (2020)	Examines how group composition impact articles' exploration trajectory and how it, in turn, impacts quality; based on the computer science theory of space trajectory and design	Compared 121 pairs of articles on the same subject from Wikipedia and Wikia	Wikipedia article	The extent of exploration is associated with higher quality, but after reaching a threshold, it is associated with reduced quality; registered contributors are less explorative in comparison to non-registered contributors	The quality scores of articles as assigned by Wikipedia contributors (available only for the Wikipedia articles)	*Production participants*: Distinguishes between registered and unregistered participants. Does not look at registration status but make the distinction between returning and new contributors: "Contributors who return to a particular article and make repeated changes are likely to develop psychological ownership (Pierce et al. 2003) and exhibit high motivation with regards to the shaping of the article" – hypothesizes on the role of veteran participants
Blumenstock (2008)	N/A	The comparison of word counts in Wikipedia's featured articles and a random sample of other articles	Wikipedia article word count	Positive association between word count and the designation of featured article status	Featured article status as an indication of quality	N/A

Chesney (2006)	N/A (deductive hypothesis development)	A survey-based assessment of the credibility of 30 articles by experts, 24 random-assignment to non-experts and Wikipedia as a whole	The perception of the credibility of the Wikipedia article and editor	Experts found articles to be more credible than non-experts; 13% of articles were reported for mistakes	Research staff assess articles' credibility, the credibility of its author, and the credibility of Wikipedia as a whole	N/A
Clauson et al. (2008)	Examines how Wikipedia articles on drugs compare in quality to professionally produced drug information database	Comparing Wikipedia entries and Medscape Drug References (MDR)	Information quality	Wikipedia entries are found to be of lesser quality across all major information quality categories	Manually evaluated using the manufacturer information and other authoritative sources across the parameters of information quality: scope, completeness, and accuracy	N/A
Dang and Ignat (2016)	Methods paper describing the use of deep learning technique for assessing Wikipedia article quality	Comparing the deep learning algorithm quality assessment of 30,000 articles to the quality scores assigned by the Wikipedia community	Article	The deep learning quality assessment found to be closely aligned with the article's prior quality scores	The quality of content as determined by the deep learning tool	N/A
De la Calzada and Dekhtyar (2010)	N/A (exploratory hypothesis development)	A comparison of two models of quality estimation of 50 stable and 29 controversial Wikipedia articles	Article quality as assessed by students	The quality of articles of different type should be computed using different means; methodological insights on using multiple amateur quality evaluators as opposed to individual experts	Article quality assessed through triangulation of quality scores made by 18-20 students	N/A
Geiger et al. (2013)	N/A	Vandalism fighting bot ClueBot NG was down at times, which is used as experimental condition for studying quality control in Wikipedia	Reverts are the predominant mechanism used to remove undesirable material, notably vandalism and spam, from Wikipedia	The overall time-to-revert was almost doubled when ClueBot NG was down; other agents did the work but much slower.	Time-to-revert defined as the amount of time between when a given edit is made to a page and when the edit is entirely removed from the encyclopedia, taking the article back to its previous state	N/A
Giles (2005)	N/A	Blind comparison by experts of 42 science articles in Wikipedia and Encyclopedia Britannica	Wikipedia articles	Articles from both sources are found to contain a similar number of errors	Quality assessment by experts (not clear on what criteria)	N/A
Greenstein and Zhu (2012)	N/A	Measures slant at a point in time over ten years, and documents its evolution over time, taking an	28,382 Wikipedia articles on US political topics	The findings show that Wikipedia contains bias, and the level or direction of bias is not fixed over	Slant as a measure of bias in articles as deviation from the neutral point of view (NPOV) rule	N/A

		approach in line with the literature examining content bias in media		time; most articles change only mildly from their initial slant		
Greenstein and Zhu (2016)	Linus's law: given enough eyeballs, all bugs are shallow	Regression analysis	19,901 Wikipedia articles on US politics	Most articles change only mildly from their initial slant (as a deviation from NPOV rule)	A unidimensional slant index (Democrat, Republican) for measuring neutrality bias	N/A
Halfaker et al. (2011)	Collective effort model (no hypotheses)	The examination of the dropout rate of editors in two sets of 200K edits, reverted and non-reverted	Wikipedia edits; edits that remain untouched longer used as a proxy for quality of edit	Reverts discourage contributions, especially reverts made by veteran providers of new providers; editors who continue to do work in Wikipedia after being reverted increase the quality of their work	Quality measured in two ways: 1) reverts/revision (the proportion of an editor's revisions that have been reverted in a given timespan), 2) PWR/word (the average number of revisions that words added by an editor persist)	*Production participants*: Reverts often target new editors instead of experienced editors; the demotivating effects concerns particularly the former. Reverts from anonymous editors are less demotivating *Production organization*: Communication about article content takes place on its talk page *Content management*: Reverts improve overall quality as editors learn from being reverted. Editors are less likely to be reverted in the future
Halfaker and Riedl (2012)	N/A	Not an empirical study	Not an empirical study	Bots are an inherent part of the Wikipedia's editorial and social system	Not an empirical study	*Production participants*: Experienced editors could not keep up with the growth of incoming contributions around 2004 after which the system had to resort increasingly to bots and automated tools to manage the system. *Production organization*: Bot Approvals Group oversees the use of bots as they can be also disruptive – bots interact directly with human editors. Bots and automated tools have changed Wikipedia culture. *Conduct management:* Bots and automated tools are most effective in combating vandalism. However, bots may be held to even higher standard than human editors.

Hu et al. (2007)	N/A (mathematical modeling)	Comparing the performance of measurement models through experiments using 242 Wikipedia articles about various countries	Article quality as indicated by its links	Authors with more authority produce higher quality edits; the length of an article is associated with its quality.	Quality is measured as the function of author authority; content that survive longer is seen as higher quality.	*Production participants*: The background of participants varies hugely. Contributor authority is related to contribution quality. *Production organization*: The article notes that Wikipedia articles are always evolving and hence their quality too. *Content management*: The literature review lists various attempts to measure Wikipedia article quality; internal article quality grading by contributors
Javanmardi and Lopes (2010)	N/A	The comparison of the time featured and non-featured articles are in a high-quality state; conceptualizes revisions that last longer as high quality	The quality of featured and non-featured articles	Average article quality increases with more edits; featured articles are of high quality 86% of the time, while non-featured are high quality 74% of the time	An author making a revision and/or revert as a proxy for revision quality; quality fluctuates over time. The quality measured based on reputation cannot asses the quality of newcomers to the system for whom there are not enough data.	*Production participants*: Experts and vandals, at the extremes, shape the quality of an article. The reputation of a user tells about the likelihood of him/her producing a high-quality contribution. *Content management*: Problems with John Seigenthaler hoax remained in Wikipedia for 132 days; article quality grading by contributors; literature review contains several papers on article quality measurement
Kane (2011)	N/A (deductive hypothesis development partially based on case study findings)	A qualitative case study of editing a featured article; a quantitative analysis of factors associated with the promotion of 188 articles nominated to receive featured article status	Wikipedia article promoted to receive a featured article status; Virginia Tech shooting article (editing case study)	The number of edits and editors is not associated with article promotion; the number of anonymous edits and breadth of top contributor experience negatively associated with article promotion; revisions and knowledge-depth of top contributors positively associated with article promotion	"Featured article" status as the indication of quality: a binary indicator of whether a nominated article was promoted to featured article status (1) or whether it was rejected (0)	*Production participants*: Anonymous contributors; committed, high-volume contributors vs. peripheral contributors. The composition of contributor community is important. *Production organization*: The article studies the dimensions of effective collaboration process in producing high-quality articles; e.g., experienced contributors (deep experience) are important in leading high quality content production. *Conduct management:* Minor forms of vandalism are not detected by bots; newcomers take time to understand various rules.

						Content management: The number of anonymous contributors is negatively related to content quality.
Kane and Ransbotham (2016)	Integrates two previous approaches to studying information quality in Wikpedia: content and collaboration	Affiliate network analysis that integrates both contributors and content; regression analysis	16,244 Medicine WikiProject articles	Wikipedia article quality ratings are found to be relatively reliable; article quality is related to affiliate network structure	An expert evaluation of quality compared to various crowdsourced measures	*Production organization*: Makes a difference between content and collaboration approaches to assessing article quality. The number of contributors and their network centrality affects the quality of articles. More contributors on an article increase quality up to a point and then decreases it. *Conduct management:* Conflict in collaboration reduces article quality.
Kittur and Kraut (2008)	Theories of group coordination (no hypotheses)	Quantitative analysis of variables associated with 23,619 articles with changing quality scores	Wikipedia article quality score	A high concentration of the main work, with more contributors in supporting roles is associated with change in quality score; equal work distribution and coordination on the talk page is not associated with quality scores.	Article quality scores assigned to a set of articles by Wikipedia volunteer administrators ranging from 'stub' to 'featured article'	*Production participants*: Core vs. peripheral participants *Production organization*: Explicit coordination is particularly important in the early stages of article development; small number of editors making the majority of work is an indication of implicit coordination. Uses Gini to measure concentration. *Content management*: Adding more editors improves quality only when they use adequate communication methods to coordinate their work. If editors don't communicate, it is better to concentrate work on an article to fewer editors.
Kriplean et al. (2007)	N/A	A purposive sampling of critical sections in talk page discussions with grounded theory analysis	The use of policies to achieve consensus on talk pages in Wikipedia	Policies help but do not solve content disputes as such as users get involved in politics and power plays.	N/A	*Production participants*: Contributors have different ideas, perspectives, and knowledge *Production organization*: Consensus seeking is a fundamental principle in Wikipedia work; contributors draw upon policies to back up their views in talk page discussions. Anyone can participate but regulars typically dominate discussions.

						Conduct management: Three-revert rule cut reverts down considerably; develop a measure for 'policy-laden' discussion, that is, policy density *Content management*: Develops a measure for 'policy-laden' discussion, that is, policy density
Lewandowski and Spree (2010)	Partially a methods paper developing a heuristic based 14 evaluation criteria to be used by the evaluators of Wikipedia article quality	Students analyzed a sample of 43 articles using the evaluation criteria	Article	Articles quality scored, on average, around 70% according to the evaluation criteria	Quality as a score based on the developed evaluation criteria	*Production organization*: Mentions tags used to indicate articles needing cleanup *Content management*: The paper notes the volatility of content as a potential issue for information quality in Wikipedia and lists numerous quality criteria for an encyclopedia. Article authority is related to its stability
Lewoniewski et al. (2016)	Uses 85 quality evaluation criteria and text mining to compare the quality and importance of Wikipedia articles in different languages	Text mining of 19,200 articles with assigned quality scores	Article	The importance of article affects its quality	Quality as a score based on 85 evaluation parameters	N/A
Lin and Wang (2019)	N/A	Regression analysis	Featured articles in Chinese Wikipedia	The results show that an increase in the number of participants increases the possibility of either enhancing or reducing the article quality. In most cases, the greater the proportion of core members, the higher the possibility of enhancing the article quality, whereas occasional participants' editorial behavior hinders quality promotion.	Wikipedia article internal quality classification	*Production participants*: Core vs. peripheral participants is an important distinction. *Production organization*: Different types of participants have different impact at different times of article development.

Liu and Ram (2011)	Input-process-output model	Descriptive statistics (cluster analysis), multiple regression	Wikipedia articles classified into different quality grades	The quality of collaboration between contributors affects article quality; high-quality articles have a core contributor or lead editors.	Contributor roles; Wikipedia article quality ratings	*Production participants*: Contributor roles and contribution patterns matter to quality; high-quality articles tend to have leaders or core contributors making the majority of the contributions. *Production organization*: Both the content of contributions and collaboration matters with respect to quality.
Salutari et al. (2020)	N/A	Combines a large-scale survey with data on Wikipedia article features to estimate the quality of user experience	Wikipedia article experience	Metrics that are generally considered as state-of-the-art for web quality of user experience do not work well with Wikipedia articles. It is difficult to predict user satisfaction.	User satisfaction	N/A
Sheppard et al. (2014)	N/A	Qualitative data from citizen science projects	Data management workflow	Develops a data model for field-based monitoring of phenomena	N/A	N/A
Stvilia et al. (2008)	N/A (grounded approach)	Case study using qualitative and quantitative techniques	Wikipedia editing activities across article, talk, policy, and user pages	When editors modify each other's contributions, it implicitly or explicitly involves an evaluation of the quality of those contributions; self-selection allows significant savings in terms of selection and coordination costs.	N/A	*Production participants*: Identify different roles in quality management, editors vs. information quality agents in particular (and vandals); anonymous users, registered users, and administrators *Production organization*: Includes a broad description and discussion of Wikipedia information quality assurance arrangements. Information quality assurance is integrated into the editorial work and focuses more on quick recovery rather than the avoidance of problems. *Content management*: Featured article process is key in promoting content (also in many other studies). Also, deletion policies are important. The paper is focused on editorial discussions in identifying content problems.

Wang and Li (2020)	N/A	A review and comparison of different computational approaches to assessing Wikipedia article quality and feature selection	Wikipedia article	Identifies three features that are most predictive of article quality	Classification to high- and low-quality articles based on Wikipedia internal classification scheme	*Production organization*: Average edits per user predicts quality among other features
Wilkinson and Huberman (2007)	N/A (no hypotheses)	Comparing editing activities in featured and non-featured Wikipedia articles	Wikipedia featured article status as a proxy for quality	Correlation found between the number of edits and distinct editors and quality	"Featured article" status as an indicator of quality	*Production organization*: The intensity of cooperative behavior increases article quality; small number of articles attract large proportion of edits and editors; quick succession of edits from different editors indicate cooperation. *Content management*: The number of edits and editors increases article quality; articles do not reach steady state but continue evolving. Uses Featured article status as the indication of quality.
Wöhner and Peters (2009)	N/A (no hypotheses)	An algorithm-based comparison of the life cycles of 100 featured and good articles and 100 articles nominated for deletion	Wikipedia articles	Identifies 11 life-cycle metrics associated with quality used for the algorithm; the length of articles and high intensity in contribution activity are associated with quality score	Article quality scores assigned to a set of articles by Wikipedia volunteer administrators ranging from 'stub' to 'featured article' status	*Content management*: The paper measures persistent and transient contributions in featured article process and article deletion. A transient contribution refers to the number of words which were changed and reversed in the same given period of time. The persistent contribution refers to all effective edits which remain in the article beyond the period.

References

Anderka M, Stein B, Lipka N (2012) Predicting quality flaws in user-generated content: The case of Wikipedia. *Proceedings of the 35th International ACM-SIGIR Conference on Research and Development in Information Retrieval* (Association for Computing Machinery, New York, NY), 981–990.

Anthony D, Smith SW, Williamson T (2009) Reputation and reliability in collective goods: The case of the online encyclopedia Wikipedia. *Rationality and Society* 21(3):283–306.

Arazy O, Kopak R (2011) On the measurability of information quality. *Journal for the American Society for Information Science and Technology* 61(1):89–99.

Arazy O, Nov O (2010) Determinants of Wikipedia quality: The roles of global and local contribution inequality. *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (Association for Computing Machinery, New York, NY), 233–236.

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.

Arazy O, Lindberg A, Rezaei M, Samorani M (2020) The evolutionary trajectories of peer-produced artifacts: Group composition, the trajectories' exploration, and the quality of artifacts. *MIS Quarterly* 44(4):2013–2053.

Blumenstock, JE (2008) Automatically assessing the quality of Wikipedia articles. UCB iSchool Report 2008-021, UC Berkeley School of Information.

Chesney T (2006) An empirical examination of Wikipedia's credibility. *First Monday* 11(11).

Clauson P, Boulos D (2008) Scope, completeness, and accuracy of drug information in Wikipedia. *The Annals of Pharmacotherapy* 42(12):1814–1821.

Dang QV, Ignat CL (2016) Quality assessment of Wikipedia articles without feature engineering. *2016 IEEE/ACM Joint Conference on Digital Libraries* (Association for Computing Machinery, New York, NY)*,* 27–30.

De la Calzada G, Dekhtyar A (2010) On measuring the quality of Wikipedia articles. *Proceedings of the 4th Workshop on Information Credibility* (Association for Computing Machinery, New York, NY), 11–18.

Geiger RS, Halfaker A (2013) When the levee breaks: Without bots, what happens to Wikipedia's quality control processes? *Proceedings of the 9th International Symposium on Open Collaboration* (Association for Computing Machinery, New York, NY), 1–6.

Giles J (2005) Internet encyclopedias go head to head. *Nature* 438(7070):900–901.

Greenstein S, Zhu F (2012) Is Wikipedia biased? *American Economic Review* 102(3):343–348.

Greenstein, S, Zhu F (2016) Open content, Linus' law, and neutral point of view. *Information Systems Research* 27(3):618–635.

Halfaker A, Kittur A, Riedl J (2011) Don't bite the newbies: How reverts affect the quantity and quality of Wikipedia work. *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (Association for Computing Machinery, New York, NY), 163–172.

Halfaker, A, Riedl J (2012) Bots and cyborgs: Wikipedia's immune system. *Computer*, 2012(March):79–82.

Hu M, Lim EP, Sun A, Lauw HW, Vuong BQ (2007) Measuring article quality in Wikipedia: Models and evaluation. *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management* (Association for Computing Machinery, New York, NY), 243–252.

Javanmardi S, Lopes C (2010) Statistical measure of quality in Wikipedia. *Proceedings of the First Workshop on Social Media Analytics* (Association for Computing Machinery, New York, NY), 132–138.

Kane GC (2011) A multimethod study of information quality in Wiki collaboration. *ACM Transactions on Management Information Systems* 2(1):article 4.

Kane GC, Ransbotham S (2016) Research note—Content and collaboration: An affiliation network approach to information quality in online peer production communities. *Information Systems Research* 27(2):424–439.

Kittur A, Kraut RE (2008) Harnessing the wisdom of crowds in Wikipedia: Quality through coordination. *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (Association for Computing Machinery, New York, NY), 37–46.

Kriplean T, Beschastnikh I, McDonald DW, Golder SA (2007) Community, consensus, coercion, control: CS*W or how policy mediates mass participation. *Proceedings of the 2007 ACM International Conference on Supporting Group Work* (Association for Computing Machinery, New York, NY), 167–176.

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.

Lewoniewski W, Węcel K, Abramowicz W (2016) Quality and importance of Wikipedia articles in different languages. Dregvaite G, Damasevicius R, eds. *Information and Software Technologies. ICIST 2016. Communications in Computer and Information Science* 639 (Springer, Berlin, Germany).

Lin Y, Wang C (2019) Wisdom of crowds: The effect of participant composition and contribution behavior on Wikipedia article quality. *Journal of Knowledge Management* 24(2):324–345.

Liu J, Ram S (2011) Who does what: Collaboration patterns in the Wikipedia and their impact on article quality. *ACM Transactions on Management Information Systems* 2(2):11–22.

Salutari F, Da Hora D, Dubuc G, Rossi D (2020) Analyzing Wikipedia users' perceived quality of experience: A large-scale study. *IEEE Transactions on Network and Service Management* 17(2):1082–1095.

Sheppard SA, Wiggins A, Terveen L (2014) Capturing quality: Retaining provenance for curated volunteer monitoring data. *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (Association for Computing Machinery, New York, NY), 1234–1245.

Stvilia B, Twidale MB, Smith LC, Gasser L (2008) Information quality work organization in Wikipedia. *Journal for the American Society for Information Science and Technology* 59(6):983–1001.

Wang P, Li X (2020) Assessing the quality of information on Wikipedia: A deep-learning approach. *Journal of the Association for Information Science and Technology* 71(1):16–28.

Wilkinson DM, Huberman BA (2007) Cooperation and quality in Wikipedia. *Proceedings of the 2007 International Symposium on Wikis* (Association for Computing Machinery, New York, NY), 157–164.

Wöhner T, Peters R (2009) Assessing the quality of Wikipedia articles with lifecycle-based metrics. *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (Association for Computing Machinery, New York, NY), article 16.

EC2 Dataset Construction

We downloaded a data dump that contains all edits to the English Wikipedia from the inception of the system on January 15, 2001 to January 1, 2017 and transformed the dump into a tabulated dataset following a similar approach applied in Aaltonen and Seiler (2016). According to 'enwiki-20170101-site_stats.sql' file in the data dump, the raw data contains the full text of 867,033,963 revisions to 41,099,459 pages since January 2001. Of these, 5,321,706 represent encyclopedia articles, while the rest of the pages are lists, redirects, portal pages and other means of organizing the content, user profiles, help pages and pages that support administrative and editorial functions such as Wikipedia policies and guidelines, categories, templates, and more. Each page may be accompanied by a talk page that allows users to discuss its content.

The downloaded data contains XML-formatted page records that encapsulate revision records, that is, individual edits to the page. The page record contains a header section with the page title, namespace, id and some optional tags such as redirect in the example below. The header section is followed by revision records. Each revision record contains an id, timestamp, contributor record, the full text of the revision and some optional tags such as comment that is a summary of the edit (not to be confused with talk page comments). Below is an example of an article page record. Note that we have truncated the record to show only the first revision of the page and slightly altered the indentation to make it easy to observe the XML structure.

```
<page>
 <title>AccessibleComputing</title>
 <ns>0</ns>
 <id>10</id>
 <redirect title="Computer accessibility" />
 <revision>
  <id>233192</id>
  <timestamp>2001-01-21T02:12:21Z</timestamp>
  <contributor>
   <username>RoseParks</username>
   <id>99</id>
  </contributor>
  <comment>*</comment>
  <model>Wikitext</model>
  <format>text/x-Wiki</format>
  <text xml:space="preserve">This subject covers
   * AssistiveTechnology
   * AccessibleSoftware
   * AccessibleWeb
   * LegalIssuesInAccessibleComputing
  </text>
  <sha1>8kul9tlwjm9oxgvqzbwuegt9b2830vw</sha1>
 </revision>
 …
<page>
```

We took a random sample of 340,000 articles and associated 172,576 talk pages (not all articles have a talk page) and transformed their XML records into an initial tabulated dataset using a purpose-built Python script. Due to a technical limitation of our computational environment, we could not process three very large articles consisting of more than 2 gigabytes of content in our sample. These are “Barack Obama” (page id 534366), “Syrian Civil War” (page id 30741795), and “Adolf Hitler” (page id 2731583). We drop these three pages from further processing. The transformation process stores page title and id, revision id, timestamp, user id and name, and other information for each sampled page revision. Each revision is represented by a row in the initial dataset. During the initial transformation, we calculated further metrics such as the Levenshtein edit distance between consecutive page revisions to estimate the amount of content change or effort put into each revision. We also analyzed the content of revisions by matching regular expression patterns against the revision content, creating a variable for each matched pattern. We then identified 4,924 articles with the presence of a cleanup template at some point in their history. We further removed pages with intractable revert patterns during quality problem events to arrive at a sample of 4,679 articles and 5,487 quality problem events as some articles contain more than one event, that is, a cleanup template appears on them more than once.

References

Aaltonen A, Seiler S (2016) Cumulative growth in user-generated content production: Evidence from Wikipedia. *Management Science* 62(7):2054–2069.

EC3 Correlation Matrix and Summary Statistics

Table EC-2. Correlation Matrix and Summary Statistics

	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11	Var12	Var13	Var14
ExperiencedOnArticle (Var1)	1													
InexperiencedOnArticle (Var2)	0.61*	1												
Policies (Var3)	0.49*	0.65*	1											
Bots (Var4)	0.43*	0.73*	0.42*	1										
ExperiencedOnTalk (Var5)	0.65*	0.47*	0.38*	0.34*	1									
InexperiencedOnTalk (Var6)	0.50*	0.73*	0.44*	0.58*	0.52*	1								
ArticleAge (Var7)	0.37*	0.21*	0.14*	0.15*	0.23*	0.11*	1							
ArticleLengthPreProblem (Var8)	0.49*	0.23*	0.18*	0.14*	0.42*	0.19*	0.41*	1						
ArticleReferencesPreProblem (Var9)	0.31*	0.12*	0.12*	0.07*	0.29*	0.11*	0.29*	0.66*	1					
ContributorsPreProblem (Var10)	0.56*	0.26*	0.27*	0.15*	0.43*	0.17*	0.50*	0.61*	0.48*	1				
ContributorsEditDist (Var11)	0.49*	0.56*	0.32*	0.37*	0.46*	0.52*	0.15*	0.42*	0.27*	0.22*	1			
BotsEditDist (Var12)	0.16*	0.25*	0.14*	0.31*	0.15*	0.24*	0.08*	0.13*	0.11*	0.08*	0.22*	1		
BotsToContributorsCountRatio (Var13)	-0.18*	-0.12*	-0.07*	0.14*	-0.12*	-0.08*	-0.09*	-0.13*	-0.06*	-0.11*	-0.12*	0.04*	1	
BotsToContributorsEditRatio (Var14)	0.40*	0.49*	0.26*	0.01	0.30*	0.34*	0.23*	0.24*	0.11*	0.23*	0.33*	0.02	-0.37*	1
Summary Statistics														
Mean	2.12	11.05	1.90	0.30	0.38	1.66	890.83	11,280	5.60	46.93	9,311	0.61	0.36	153.84
Std. dev.	2.78	19.71	2.48	1.51	1.02	3.11	1005.97	16,781	19.48	113.95	22,460	0.20	0.18	963.82
Min.	0	0	0	0	0	0	0	0	0	0	2	0.001	0	-18,171
Max.	42	370	33	31	17	83	5,353	207,762	485	2,160	502,958	1	0.85	18,397

EC4 Assessing Proportional Hazards Assumption

One of the main assumptions of the Cox proportional hazard model is proportionality. There are several methods for verifying that a model satisfies the assumption of proportionality. The first is to include interaction terms between the time-varying independent variables and the logarithm of the cleanup time (Hosmer et al. 2008), with the effects not changing over time, except in ways already accounted for in the model where the dependent variable is CleanupTime. The estimated coefficients of these interaction terms should not be significantly different from zero under the null hypothesis. Thus, following Cleves et al. (2016) and Hosmer et al. (2008), we interacted all independent variables of interest with the logarithm of time and estimated a basic Cox PH model that includes those interaction terms. If a time-dependent covariate is significant this indicates a violation of the proportionality assumption for that specific predictor. For all the estimated coefficients of the interaction items, we fail to reject the null hypothesis indicating that the proportionality assumption is not supported.

Table EC-3. Coefficients of the Time-Dependent Covariates

	Coefficient
ExperiencedOnArticle	0.1374***
	(0.0259)
InexperiencedOnArticle	-0.1052***
	(0.0120)
PolicyCitations	0.3007***
	(0.0651)
Bots	-1.0825***
	(0.0722)

Notes. * p<0.1; ** p<0.05; *** p<0.01

The second method to check the proportionality of hazards assumption is based on analyzing the regression residuals from a model with CleanupTime as the dependent variable (Cleves et al. 2016; Grambsch and Therneau 1994). The intuition is to "retrieve the residuals, fit a smooth function of time to them, and test whether there is a relationship" (Cleves et al. 2016, p. 200). It includes both a global test on all the independent variables and tests on individual independent variables. Our results reject the proportionality assumption in both the global test and the tests on individual independent variables of interest. Violating the proportional-hazard assumption means that the influence of these individual variables might not be constant over time. These tests again indicate that the proportionality assumption should be rejected.

Table EC-4. Residuals χ^2 Estimates on Individual Variables and a Global Test			
	rho	**χ**	**Prob>χ^2**
ExperiencedOnArticle	-0.04886	9.61	0.0019
InexperiencedOnArticle	-0.05102	16.89	0.0000
PolicyCitations	-0.01105	0.62	0.4294
Bots	0.20742	282.16	0.0000
ExperiencedOnArticle # Policies	-0.00322	0.05	0.8166
InexperiencedOnArticle # Policies	0.02592	5.13	0.0235
ExperiencedOnArticle # Bots	0.01414	0.87	0.3513
InexperiencedOnArticle # Bots	-0.04315	14.81	0.0001
ExperiencedOnTalk	-0.06315	14.85	0.0001
InexperiencedOnTalk	0.05492	14.64	0.0001
ArticleAge	0.11436	47.93	0.0000
ArticleLengthPreProblem	-0.04665	4.74	0.0295
ArticleReferencesPreProblem	0.01563	1.61	0.2042
ContributorsPreProblem	0.00373	0.05	0.8185
ContributorsEditDist	0.01019	0.33	0.5671
BotsEditDist	-0.00591	1.03	0.3102
BotsToContributorsCountRatio	0.04349	6.33	0.0119
BotsToContributorsEditRatio	0.13341	123.67	0.0000
Global test		514.23	0.0000

To summarize, the survival of different individuals is different in our data, and the assumption of homogeneity of individuals in the population studied is not well-founded due to, among others, the existence of unknown or unobserved risk factors that are not included in the model.

References

Cleves M, Gould W, Gutierrez R, Marchenko Y (2016) *An Introduction to Survival Analysis Using Stata* (Stata press).

Grambsch PM, Therneau TM (1994) Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 81(3):515–526.

Hosmer DW, Lemeshow S, May S (2008) *Applied Survival Analysis: Regression Modeling of Time to Event Data* 2nd ed. (John Wiley, New York, NY).

EC5 Variance-Covariance Matrix of the Main Model

Table EC-5. Model Variance Covariance Matrix

	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11	Var12	Var13	Var14	Var15	Var16	Var17	Var18
ExperiencedOnArticle (Var1)	1																	
InexperiencedOnArticle (Var2)	-0.27	1																
Policies (Var3)	0.16	-0.36	1															
Bots (Var4)	0.02	-0.70	-0.01	1														
ExperiencedOnArticle # Policies (Var5)	-0.23	0.04	-0.67	0.18	1													
InexperiencedOnArticle # Policies (Var6)	-0.16	0.60	-0.46	-0.37	-0.14	1												
ExperiencedOnArticle # Bots (Var7)	-0.77	0.22	-0.10	-0.07	0.10	0.18	1											
InexperiencedOnArticle # Bots (Var8)	0.39	-0.73	0.43	0.41	0.03	-0.92	-0.44	1										
ExperiencedOnTalk (Var9)	0.11	-0.14	0.02	0.09	-0.06	-0.19	-0.29	0.29	1									
InexperiencedOnTalk (Var10)	-0.04	0.06	-0.01	0.11	-0.16	0.16	0.09	-0.15	-0.14	1								
ArticleAge (Var11)	-0.24	-0.01	-0.10	0.24	0.15	-0.03	0.17	-0.04	-0.04	0.25	1							
ArticleLengthPreProblem (Var12)	-0.16	0.14	-0.02	-0.23	0.06	0.02	0.14	-0.06	-0.15	0.04	-0.16	1						
ArticleReferencesPreProblem (Var13)	-0.02	-0.17	0.18	0.26	-0.04	-0.30	0.07	0.24	-0.06	0.07	0.14	-0.41	1					
ContributorsPreProblem (Var14)	-0.04	0.05	-0.10	-0.09	-0.10	0.25	-0.22	-0.14	0.05	-0.04	-0.19	-0.23	-0.40	1				
ContributorsEditDist (Var15)	-0.04	-0.13	-0.02	0.24	0.01	-0.06	-0.03	0.09	-0.07	-0.11	0.08	-0.20	0.04	0.05	1			
BotsEditDist (Var16)	0.13	0.24	-0.07	-0.51	-0.15	0.52	-0.17	-0.44	-0.19	-0.34	-0.27	0.05	-0.36	0.31	-0.34	1		
BotsToContributorsCountRatio (Var17)	0.16	-0.08	0.05	0.11	-0.09	-0.06	-0.10	0.10	0.06	0.14	0.09	0.03	-0.01	-0.03	0.02	-0.14	1	
BotsToContributorsEditRatio (Var18)	0.05	-0.74	0.10	0.67	0.03	-0.26	-0.05	0.35	0.06	-0.02	0.03	-0.17	0.09	0.02	0.03	-0.11	0.31	1
lnsigma	0.02	-0.25	0.04	0.47	0.02	-0.18	-0.03	0.17	0.06	0.10	0.24	-0.17	0.13	-0.09	0.10	-0.28	0.00	0.28
kappa	0.04	0.18	0.01	-0.43	-0.06	0.11	-0.04	-0.08	0.01	-0.13	-0.29	0.22	-0.16	0.13	-0.04	0.24	-0.02	-0.25

EC6 Choice of Distribution for the Survival Time Model

We fit a parametric random-effects survival-time model. For estimation purposes, we fit a generalized gamma distribution for the conditional distribution of the response given the random effects. We made this choice after comparing different distributions and selecting the one that resulted in the lowest Akaike's Information Criterion (AIC). We use robust standard error estimation. We also use the Huber/White/sandwich variance-covariance estimator (Arellano 2003, Wooldridge 2010).

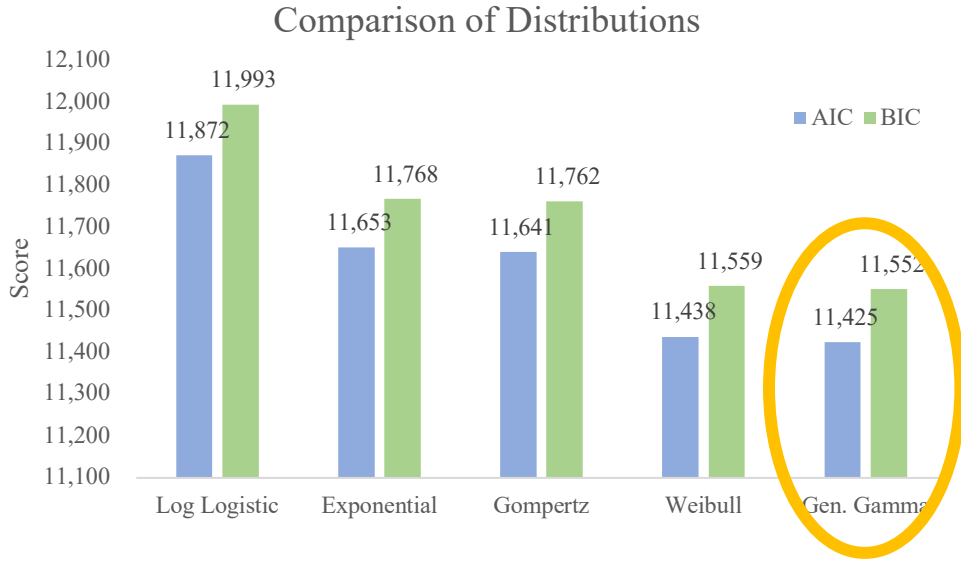

Figure EC-1. Comparison of Performance for Survival Time Model using various Distributions

References

Arellano M (2003) *Panel Data Econometrics* (Oxford University Press, Oxford, UK).
Wooldridge JM (2010) *Econometric Analysis of Cross Section and Panel Data* (The MIT Press, Cambridge, MA).

EC7 Analysis with and Without Stub Articles

The main analysis excluded articles that were declared stubs at the end of the cleanup event. We conducted robustness analysis for Equation 1 by (i) excluding articles that were stubs at the start of the cleanup event (without Start Stubs) and (ii) not excluding any stubs articles (with Stubs).

Table EC-6. Survival Analysis Results		
	Without Start Stub	**With Stubs**
Control Mechanisms		
Bots	0.3736***	0.3782***
	(0.0298)	(0.0268)
PolicyCitations	-0.2888***	-0.2732***
	(0.0385)	(0.0365)
Contributor Experience		
ExperiencedOnArticle	-0.0981***	-0.0949***
	(0.0191)	(0.0182)
InexperiencedOnArticle	0.0075	0.0064
	(0.0049)	(0.0047)
Interactions		
ExperiencedOnArticle × Bots	0.0077***	0.0077***
	(0.0029)	(0.0028)
InexperiencedOnArticle × Bots	-0.0028***	-0.0029***
	(0.0004)	(0.0004)
ExperiencedOnArticle × PolicyCitations	0.0053**	0.0044*
	(0.0025)	(0.0024)
InexperiencedOnArticle × PolicyCitations	0.0021***	0.0021***
	(0.0004)	(0.0003)
Explicit Coordination		
ExperiencedOnTalk	-0.1698***	-0.1794***
	(0.0315)	(0.0311)
InexperiencedOnTalk	0.0866***	0.0923***
	(0.0114)	(0.0105)
Article Characteristics		
ArticleAge	1.31E-04***	1.46E-04***
	(2.38E-05)	(2.21E-05)
ArticleLengthPreProblem	-8.08E-06***	-8.66E-06***
	(2.32E-06)	(2.32E-06)
ArticleReferencesPreProblem	0.0042*	0.0050**
	(0.0022)	(0.0024)
Edit History		
ContributorsPreProblem	-0.0026***	-0.0027***
	(0.0003)	(0.0003)

ContributorsEditDist		-1.19E-06	-1.32E-06
		(1.44E-06)	(1.44E-06)
BotsEditDist		1.25E-04	1.34E-04
		(2.60E-04)	(2.44E-04)
BotsToContributorsCountRatio		-0.1088	-0.1026
		(0.1168)	(0.1003)
BotsToContributorsEditRatio		0.1159***	0.1142***
		(0.0126)	(0.0114)
	AIC	11,169.31	13,715.82
	BIC	11,296.18	13,847.25
	Observations	4,428	5,487

Notes. * p<0.1; ** p<0.05; *** p<0.01

EC8 Analysis with Policies Disaggregated into Conduct and Conduct Policies

Many of the most often cited Wikipedia policies fall into two categories. Conduct policies provide general guidelines on desired and undesired behavior in the form of, for instance, anti-vandalism, anti-harassment, and civility guidelines. Content policies are more specific guidelines concerning the content of an article such as on the verifiability of information, use of images, and references. We disaggregated the overall policy citations to conduct policy citations and content policy citations to assess whether the type of policy makes a difference to our results.

In the main analysis, we found that there is a marginal decrease in cleanup time that is larger when an increase in the number of contributors, with or without prior experience in editing an article, is accompanied by fewer policy citations during a quality problem event. Here, we find that citations of conduct policies *and* the citations of content policies reduce cleanup time with related larger marginal decrease in cleanup time with an increase in the number of contributors, with or without prior experience. Overall, these findings provide further support for H3. As such, when we consider a more granular treatment of policies, the results reinforce the finding that it is the inexperienced contributors who are likely to be slowed down, while the experienced contributors are unlikely to be slowed down in aiding with the cleanup process.

Table EC-7. Survival Analysis Results for the Model with Disaggregated Policies and their interactions

	Coefficient
Control Mechanisms	
Bots	0.3855***
	(0.0322)
ContentPolicies	-0.2556***
	(0.0677)
ConductPolicies	-0.333***
	(0.0547)
Contributor Experience	
ExperiencedOnArticle	-0.0907***
	(0.0195)
InexperiencedOnArticle	0.0094
	(0.0058)
Interactions	
ExperiencedOnArticle × Bots	0.0073***
	(0.0028)
InexperiencedOnArticle × Bots	-0.0031***
	(5.74E-04)
ExperiencedOnArticle × ContentPolicies	0.0001
	(0.0057)
InexperiencedOnArticle × ContentPolicies	0.0036**

	(0.0018)
ExperiencedOnArticle × ConductPolicies	0.0056
	(0.0037)
InexperiencedOnArticle × ConductPolicies	0.0023***
	(0.0005)
Explicit Coordination	
ExperiencedOnTalk	-0.171***
	(0.0313)
InexperiencedOnTalk	0.0755***
	(0.0119)
Article Characteristics	
ArticleAge	1.25E-04***
	(2.40E-05)
ArticleLengthPreProblem	-8.15E-06***
	(2.33E-06)
ArticleReferencesPreProblem	0.0038*
	(0.0022)
Edit History	
ContributorsPreProblem	-0.0025***
	(0.0003)
ContributorsEditDist	-1.68E-06
	(1.46E-06)
BotsEditDist	8.56E-05
	(2.57E-04)
BotsToContributorsCountRatio	-0.1171
	(0.1165)
BotsToContributorsEditRatio	0.1147***
	(0.0131)
AIC	11,168.81
BIC	11,313.81
Observations	4,473

Notes. * p<0.1; ** p<0.05; *** p<0.01