

# A Multimethod Study of Information Quality in Wiki Collaboration

GERALD C. KANE, Boston College

In this article, the author presents the results of a two-phase, multimethod study of wiki-based collaboration in an attempt to better understand how peer-produced collaboration is done well in wiki environments. Phase 1 involves an in-depth case study of the collaborative processes surrounding the development of the Wikipedia article on the 2007 Virginia Tech massacre. The rich data collected are used to develop an initial set of testable hypotheses of factors that enhance the quality of peer-produced information in wiki environments. Phase 2 tests these theories through a quantitative analysis of the collaborative features associated with 188 similar articles that Wikipedia considered for recognition as their best (i.e., the top 0.1%). Four collaborative features are examined for their effects on quality: volume of contributor activity, type of contributor activity, number of anonymous contributors, and top contributor experience. Volume of contributor activity is the only feature that is unsupported, a particularly interesting result because previous literature connects that factor most clearly to success in wiki-based collaboration. Implications are discussed.

Categories and Subject Descriptors: H.1.2 [Models and Principles]: User/Machine Systems

General Terms: Human Factors, Performance, Management, Theory

Additional Key Words and Phrases: Wiki, collaboration, electronic collaboration, virtual teams, electronic communities, wiki, multimethod studies, Wikipedia, Web 2.0, anonymity, shaping, information quality, peer-production.

## ACM Reference Format:

Kane, G. C. 2011. A multimethod study of information quality in wiki collaboration. *ACM Trans. Manag. Inform. Syst.* 2, 1, Article 4 (March 2011), 16 pages.  
DOI = 10.1145/1929916.1929920 <http://doi.acm.org/10.1145/1929916.1929920>

## 1. INTRODUCTION

For decades, companies have been interested in the promise of effective knowledge management: the ability to get the right knowledge to the right people at the right time [Alavi and Leidner 2001]. Unfortunately, many efforts to develop effective knowledge management systems and processes have failed to live up to these promises, partially because employees did not take the time and effort needed to package and store their knowledge in these formal repositories [Bock et al. 2005; Griffith et al. 2003; Kankanhalli et al. 2005]. A new generation of Internet-based collaborative tools, commonly known as social media or Web 2.0, can potentially overcome many of the problems of previous knowledge management efforts by integrating employees' collaborative work more naturally. Furthermore, by residing on accessible Internet-based platforms, companies can also invite nonemployees (e.g., passionate customers, members of a practice community) to contribute knowledge and collaborate as well. As a result, companies are increasingly turning to social media-based peer production as a

---

This research is supported by the National Science Foundation, Grant No. 0953285.

Author's address: G. C. Kane, Department of Information Systems, Carroll School of Management, Boston College, Chestnut Hill, MA; email: [gerald.kane@bc.edu](mailto:gerald.kane@bc.edu).

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or [permissions@acm.org](mailto:permissions@acm.org).

© 2011 ACM 2158-656X/2011/03-ART4 \$10.00

DOI 10.1145/1929916.1929920 <http://doi.acm.org/10.1145/1929916.1929920>

valuable mechanism for generating and disseminating valuable information [Li and Bernoff 2008; Tapscott and Williams 2006].

A wiki, a Web site that users can collaboratively edit, is a social media platform widely adopted for peer production. Companies are increasingly employing wikis as platforms for electronic collaboration [Kane and Fichman 2009; Wagner and Majchrzak 2006]. Novell, for example, has experimented with wikis to allow their users to collaboratively develop support, guidance, and documentation [Wagner and Majchrzak 2006]. Nonprofit organizations, such as the Cleveland Clinic, have developed wiki platforms to assemble a knowledge resource for medical professionals [Kane et al. 2009a]. The highest profile example of wiki-based collaboration is Wikipedia, which uses a wiki platform harnessing the efforts of thousands of volunteer contributors to develop an encyclopedia comprising more than 16M articles in hundreds of languages.

Despite the widespread adoption of wikis and other social media, little is understood about how users might collaborate most effectively when using these tools. Although Gartner Consulting estimates that 60% of the Fortune 1000 are experimenting with wikis and other forms of social media to support collaboration, most of these efforts will fail to generate valuable information [Sarner 2008]. Technological problems do not cause these failures; the efforts fail because their collaborative processes are ineffective. The literature has recognized that Wikipedia is a valuable environment for studying peer-production processes [Kane and Fichman 2009] and has examined it widely [Kittur and Kraut 2008; Kittur et al. 2009; Ortega et al. 2008]. Although Wikipedia represents a unique collaborative environment, recent research has suggested that these processes may also extend to collaboration in other social media settings [Kittur and Kraut 2010].

In this article, I present the results of a two-phase, multi-method study of collaboration on Wikipedia to better understand how quality information is generated in wiki-based peer-production settings. Phase 1 is an in-depth case study of the Wikipedia article about the 2007 Virginia Tech (VT) massacre, used to gather rich data for hypothesis building about the quality of peer-produced information in wiki-based environments. As one of the most highly profiled, extensively edited, heavily vandalized, and widely praised articles in Wikipedia history, this article provides rich insight into the collaborative processes that govern wiki-based collaboration and generates several testable hypotheses. The hypotheses are tested in Phase 2, a quantitative analysis of 188 Wikipedia articles that were nominated as featured articles during the time period covered by the Phase 1 analysis (April to June 2007). Wikipedia selects its featured articles for representing the highest standards of collaboration: they are stable, neutral, well-written, well-researched, and comprehensive.<sup>1</sup> As such, featured article status can be used as one measure, albeit not the only, of the quality of information generated in social media-based peer-production environments.

## 2. RESEARCH METHOD AND SETTING

Drawn from the Hawaiian word meaning *quick*, a wiki is simply a Web site that anyone can edit. Wikipedia uses a wiki platform to support an open-source encyclopedia; anyone can edit any article. Established in 2001, the English version of Wikipedia has, as of this writing, developed around 3.5M separate articles. Wikipedia supports an additional 13M articles published in hundreds of other languages. The system records changes, contributors' identities, and descriptions, including the times changes were made.

---

<sup>1</sup>[http://en.wikipedia.org/wiki/Wikipedia:Featured\\_article\\_criteria](http://en.wikipedia.org/wiki/Wikipedia:Featured_article_criteria).

Table I. Data Collected for Phase 1 Research

Data	Description
Version Comparison	Collected incremental versions of the article over time preserved by wiki platform. Collected complete versions of the article hourly for first 100 hours, daily for next 10 days, and triweekly for first 3 months. Placed versions into Excel spreadsheet for visualization and analysis, permitting longitudinal analysis of article development.
Edit Log	Compiled contributor activity preserved in the article's history tab. Aggregated data to determine behavior and structure of collaborative community. Analyzed total number of contributions (9,200), total contributors (1,700), and their distribution over time. Examined total contributions/contributor, minor contributions, and anonymous contributions to identify top 1%, 2%, and 10% of contributors. Validated data in edit logs through comparisons with actual contributions, and used these data to validate automated tools to collect collaborative data for Phase 2.
User Pages	Detailed self-reported identities of contributors and their histories. Created rough profile of top 2% of contributors, including biographical data on user pages and previous experience editing Wikipedia from archives.
Talk pages	Talk pages are forums for collaborators to discuss processes. Coded comments to determine issues faced during article development. With 2 research assistants coded approximately 4,200 comments in 625 distinct discussion threads to determine issues facing community during collaboration (90% IRR, Cohen's Kappa).
Interviews	Surveyed top 2% of contributors ( $n = 34$ , 17 usable responses) about collaborative features of VT article, their role in article development, and motivations for contributing. Followed up with email for clarification and validation or observations. Clarified responses through email or telephone follow-up.

### 3. CASE STUDY OF THE VT MASSACRE ARTICLE ON WIKIPEDIA

Because little is known about wiki-based collaboration, the author first conducted an in-depth single-case study to develop a set of testable hypotheses. A critical-case methodology was chosen [Yin 2008] because a highly successful article developed under extremely challenging conditions would be most likely to accentuate the features of wiki-based collaboration most associated with information quality. Wikipedia preserves virtually the entire record of collaborative activities that support article development, which can be triangulated to provide a remarkably rich picture of wiki-based collaboration [Kane and Fichman 2009; Miles and Huberman 1994]. Table I details these data sources. The data were collected and analyzed over three months (August to October 2007) to identify the salient aspects of collaboration that led to successful collaboration on the article examined.

The article on the VT massacre was chosen as the critical case to be examined for several reasons. First, it is a clear example of high-quality peer-produced information. It was awarded featured article status, a quality designation awarded to the best 0.1% of all Wikipedia articles. The mainstream media also praised it for its exemplary treatment of the event [Dee 2007]. Second, the article was subjected to an extremely high volume of collaborative activity in a short period. Within a week, it became one of the most heavily edited articles in Wikipedia history, with almost 8,000 contributions. It was also one of the most frequently viewed articles on Wikipedia at the time, averaging more than 135,000 hits per day in the weeks following the event and 143,000 hits on the first anniversary of the incident. Third, the high profile nature of the occurrence meant the article attracted diverse contributors. Fourth, the article was developed under extremely challenging conditions, subject to incomplete and rapidly changing information. Within days, it also became one of the most heavily vandalized articles across Wikipedia. Thus, because this case represents one of the most visible, highly

Table II. Analysis of Discussion Threads on Article's Talk Pages

Threads	Comments	Category of Comment	Category Description
138 (22%)	1028 (24%)	Content	Deciding whether to include or omit certain information (e.g., biographical entries for victims?).
142 (23%)	736 (17%)	Style	Style, formatting, and organization (e.g., picture inclusion or location, number and titles of sections).
72 (12%)	1100 (26%)	Semantics	Word choice (e.g., <i>shooting</i> or <i>massacre</i> ?).
93 (15%)	494 (12%)	Factual Accuracy	Determining whether certain information in the article was correct (e.g., number of fatalities).
72 (12%)	200 (5%)	Wikipedia Policies	Comments citing specific Wikipedia rules affecting content development (e.g., neutral point-of-view).
39 (6%)	228 (5%)	Administration	Comments focused on required roles and tasks (e.g., would someone focus on cleaning up?).
28 (4%)	299 (7%)	Offline Discussion	Comments unrelated to article development (e.g., expressions of grief).
42 (7%)	160 (4%)	References	Comments addressing need for adequate references; authority of references (e.g., student blogs).
<b>626</b>	<b>4245</b>	<b>TOTAL</b>	

praised, heavily edited, and intensely vandalized articles on Wikipedia, it is ideal for deeply examining wiki-based collaborative processes.

A team of research assistants supervised by the researcher analyzed these qualitative data in several ways. First, they collected hourly snapshots of the article content that were then placed into a spreadsheet to allow visualization of the article over time. These data showed how the article changed and evolved as collaborators developed it. Second, teams independently read the discussion pages and coded the content and collaborative activity in each discussion thread: type of content (e.g., semantics) and specific issues (e.g., whether to use the term *shooting* or *massacre*?).<sup>2</sup> These data provided insight regarding issues collaborators faced when developing information as well as their collaborative processes. Table II shows code categories used and instances identified. Discussion threads were then cross-referenced with actual article activity to see how the discussions mirrored the actual collaboration. Third, the author used data from both the discussion threads and the article history to identify the group of top contributors to the article. The researcher posted a link on their user pages to a short survey to gain their personal perspectives on the collaborative process and to learn how their collaboration on this article was similar to or different from their work on other Wikipedia articles. Usable responses came from 17 top contributors, a 50% response rate. Responses were then aggregated and coded to confirm or provide greater insight into the themes identified in the coding.

The researchers iterated between phases of data collection and analysis to better understand and refine their understanding of social media-based peer production. The remainder of this section documents how these data were used in conjunction with the previous literature to develop a series of hypotheses regarding the quality of information in peer-production settings specifically relating to: (1) contribution volume, (2) type of contributor activity, (3) role of anonymous contributors, and (4) top contributor experience.

### 3.1. Contribution Volume

Recent research on Wikipedia has found that the volume of contributors and contributor activity is the most important factor in determining the quality of

<sup>2</sup>This particular issue, choosing between the terms *shooting* and *massacre*, occupied 20% of the total comments on the discussion page.

peer-produced information [Ball 2007; Kittur et al. 2007; Wilkinson and Huberman 2007]. More contributors can help prevent errors of *commission* because they increase the likelihood that someone will identify and correct errors. A similar philosophy operates in open-source software development: “[W]ith enough eyeballs, all bugs are shallow” [Raymond 1999]. More contributors also guard against errors of *omission*. Each contributor brings different information, abilities, and mental frameworks, increasing the intellectual resources available to article development. As contributors read existing content, they compare it with their own knowledge, identify missing information, and contribute that information to the content.

The volume of contributor activity is also important to the quality of information created in the VT massacre article. One respondent said, “The most important factor [to this article’s success] was the massive amount of available manpower willing to improve the article.” He noted that more people collaborating on an article means that more people copyedit, add content, patrol for vandalism, and complete other necessary tasks. Some respondents, however, also indicated another reason for having many collaborators: Wikipedia has strong norms that govern collaborative activities [Butler et al. 2008]. One respondent indicated that more collaborators and contributor activity means greater community pressure to adhere to community norms and Wikipedia’s standards: “There was zero tolerance for unsubstantiated comments. If you didn’t back it up, it was gone. So it never contained any ‘random, speculative, heard-it-somewhere pseudo-information.’” He was quoting Jimmy Wales, Wikipedia founder, who spoke about maintaining standards (although not in reference to this article). A higher volume of contributor activity made the community rely more heavily on these norms to govern the article’s development. The first hypothesis reflects these observations:

*H1a.* The total number of contributors involved in article development will be positively related to the quality of peer-produced information.

Yet, the quantity of contributors may differ significantly from the quantity of their work. Research on online communities also suggests that most participants contribute very little to the community; a relatively few perform most of the work [Lave and Wenger 1991; Preece and Schneiderman 2009]. Thus, it is not only important to have more contributors but also that these contributors are willing to work on developing content. Previous research on Wikipedia has noted that the concentration of contributor activity is also positively related to the quality of information produced by the community [Kittur et al. 2007].

The case data supported these previous findings. Collaboration on the article demonstrated a core periphery structure in which relatively few contributors disproportionately contributed the most. The top 1% of contributors made more than 20% of the total contributions, and the top 10% of contributors contributed more than 60%. Most (69%) contributed only once or twice. Although some contributors were only peripherally involved in the article’s development, others were highly committed, devoting many hours to its development. These contributors who spent more time working on the article were more likely to make many constructive contributions and to actively participate in the article’s talk pages.

These core contributors also played another important role in the community: protecting the content from unhelpful changes by members of the periphery. Approximately 15% of the activity to the talk pages was dedicated to contributions that were reverted or undone by members of the core community. They requested justification from the contributor regarding how and why the contribution would improve the content before it would be allowed. This protecting activity further demonstrated the commitment of core members to the quality of the content: contributors who make many contributions are more likely to be highly committed; contributors who make relatively

few contributions are unlikely to be highly committed. Thus, the overall commitment of contributors to the article development is likely to be related to its quality. This led to a related hypothesis:

*H1b.* The average number of contributions per contributor will be positively related to the quality of peer-produced information.

### 3.2. Type of Contributor Activity

A theoretically distinctive behavior of wiki-based collaboration is the process of shaping or packaging the content created by others [Kane et al. 2009b; Yates et al. 2010]. Markus [2001] identified similar processes as “the process of culling, cleaning and polishing, structuring, formatting, or indexing documents against a classification scheme.” [Markus 2001, p. 60]. Shaping content is important for consumers because it presents and structures knowledge in a way that makes its value and relevance clearly identifiable. The amount of contributor activity spent shaping a Wikipedia article for general consumption is likely to be positively related to information quality, because it cultivates perceptions that the content has a professional writing style with few, if any, typographical or grammatical errors.

The case data supported the importance of these shaping processes. Although the data showed that a few individuals were responsible for most contributions, a closer look revealed that these contributors were responsible for a particular type of contribution. Surveys of the most prolific contributors indicated that they typically did more editing and organizing than contributing new content. A closer examination of the activity of the top 1% of contributors demonstrated that, indeed, they spent most of their efforts copyediting (53%) or deleting (29%) and a small percentage of their activity adding content (2%). Most of their contributions addressed style, format, or structure; an examination of the discussion threads showed top contributors were more concerned about how to present the content clearly, compellingly, and professionally than about whether content was appropriate or accurate.

This process of shaping the article content [Kane et al. 2009b; Yates et al. 2010] was also evident in the overall patterns of contributions over time. Adding and deleting activities followed a distinctive pattern. The volume of contributor contributions demonstrated that, particularly in the week following the event, contributions varied depending on U.S. time (see Figure 1); activity was highest at midday and lowest at night. At its peak, the article received 350 contributions per hour, approximately one every 10 seconds.

Observations of article changes showed that the types of contributions differed considerably during these times. During the overnight hours, contributors (often from Europe or Australia) typically made changes to the style and organization of the article: removing duplicate information, organizing content, correcting typos, and improving grammar. Respondents indicated this pattern resulted because the volume of editorial activity was simply too high during the daylight U.S. hours to make such shaping changes. As a result, successive cycles of expansion and contraction emerged; in the daytime the article expanded when contributors added new information as it became available and verifiable; during the nighttime the article contracted as contributors shaped content contributed by others into a more coherent whole.

Thus, the process of shaping the article consumed a significant part of the collaborative activity. One top contributor reported that nearly 75% of her contributions were these types of stylistic, shaping edits, a report that was validated by a random check on the edit log. The community’s attention to these stylistic details will lead to perceptions of a high-quality collaborative product [Markus 2001; Stein and Zwass 1995]. These observations and arguments led to the second hypothesis.

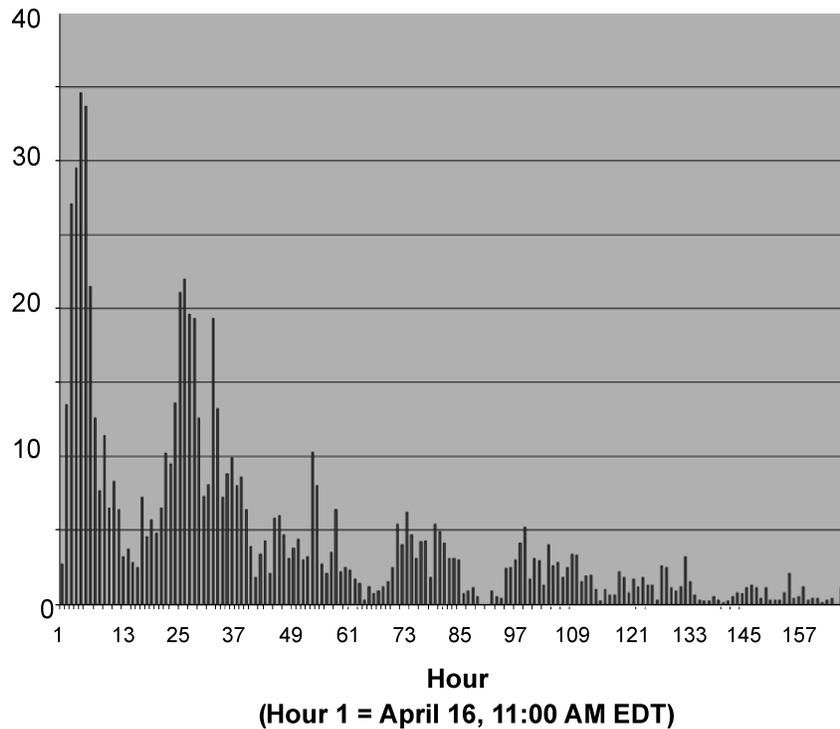


Fig. 1. Number of contributions per hour (Week 1).

*H2.* The overall level of content shaping will be positively related to the quality of peer-produced information.

### 3.3. Role of Anonymous Contributors

Researchers have studied extensively the influence of anonymity on collaboration [Sia et al. 2002]. Anonymity influences group behavior primarily because it lowers the level of social presence in a group: the degree to which people establish personal connections in a communication setting [Short et al. 1976; Sia et al. 2002]. Anonymity has also been found to increase disruptive behavior and polarizing dynamics in collaborative environments [Jessup et al. 1990; Sia et al. 2002]. Disruptive behavior, such as vandalism, can hinder information quality. Polarizing dynamics can make it more difficult for the contributor community to reach consensus, the standard for decision-making on Wikipedia. Although anonymity has been found to enhance collaboration in some studies [Connolly et al. 1990; Siegel et al. 1986; Valacich et al. 1992], these benefits are largely for the generation of novel and unique ideas. In contrast, Wikipedia expressly prohibits novel ideas and allows only information referenced from authoritative sources.

On Wikipedia, all contributors typically adopt usernames that are typically unconnected to their real-world identity, so most contributors are pseudonymous. Since Wikipedia tracks the entire contribution history of all user's accounts and allows users to post information about themselves on a user page; however, pseudonymity is not a problem for collaboration. Users can develop a history and an informal reputation. Anonymity, however, means that a particular user may not be connected to past user history and other activities in the wider Wikipedia community. Users who choose to

avoid adopting a Wikipedia identity and to log in under that identity are identified only by IP addresses.

The role of anonymous contributors was one of the biggest challenges that the contributor community faced in developing the VT massacre article. The high-profile nature of the event attracted mischief-makers who targeted the article for vandalism by deliberately tampering with the contents to undermine the article's value. An automated bot that patrolled Wikipedia for signs of vandalism indicated that the VT massacre article was the third most heavily vandalized on Wikipedia at the time (behind Germany and Cannabis). Only 90 minutes after the event, the first instance of sabotage claimed that Alberto Gonzales, then U.S. Attorney General, was responsible for the shootings. Edit logs show that anonymous contributors, identified only by IP addresses, did the most harm. Nevertheless, vandalism appeared to cause no direct problems to article development. Examining 60 cases of major vandalism (e.g., where large sections of the article were deleted or profanity/hate speech was added), the edit log showed that these disruptions were usually corrected in less than a minute by vigilant contributors or automated software bots programmed to fix the most common forms of vandalism.

Vandalism did, however, hamper article development indirectly because the contributors had to remain vigilant for minor forms of sabotage not automatically corrected by the bots. For instance, an anonymous contributor claimed to be the individual interviewed in a press report and attempted to clarify information attributed to him. The contributor community had difficulty evaluating his claims and remained skeptical primarily because he was editing anonymously. The community decided to eliminate the contribution simply because it could not be fully verified. Anonymity appeared to negatively affect article development because it indirectly connected to instances of vandalism, making it more difficult to trust, evaluate, and integrate community contributions. These observations and arguments led to a third hypothesis.

*H3* The number of anonymous contributors on an article will be negatively related to the quality of peer-produced information.

### 3.4. Top Contributor Experience

Previous research has suggested that the leaders of a community often exert considerable influence over collaboration in online communities [Lave and Wenger 1991; Preece and Schneiderman 2009]. As a result, the ability of these leaders may be related to the quality of information created by the community. One way to measure a leader's ability is through experience [Kane 2009]. The more they collaborate on the platform, the more they learn the rules, norms, and procedures for effective collaboration in the community [Kane 2009]. Some respondents indicated that they spent more than 40 hours per week editing Wikipedia, feasibly developing considerable collaborative experience in this environment. Nevertheless, since "anyone can edit" an article on Wikipedia, one cannot assume that the most experienced contributors will become the leaders of the community.

Case data suggested that the experience of key players in the contributor community was important for understanding successful collaboration on the VT massacre article. Although not formally occupying any leadership position, a single key figure tended to lead the development effort. Other community members tended to cede to the decisions and guidance of this top contributor. She noted that inexperienced but well-meaning contributors often created difficulties in the editing process.

"Those who are trying to contribute to the article productively, but doing so in a way that violates various Wikipedia rules or is unacceptable for some other reason, have to be handled with a lot more patience, politeness, and a willingness to explain basically everything. This is all very time consuming and can be

somewhat disconcerting to the new user, who has merely heard that Wikipedia is the encyclopedia that anyone can edit, and hasn't heard that there are some rules."

Interestingly, one key figure emerged to lead in the first two days of collaboration, but another emerged for the remainder of the collaboration.

The development process functioned more smoothly when experienced leaders stepped forward and inexperienced contributors were willing to take secondary roles (as evidenced by decreased activity on the talk pages), rather than trying to lead the effort. For instance, as the community's leadership was being established, one highly active but inexperienced person created a number of issues because of good faith efforts to improve the wording and language of the article that violated the community norms for making such changes. Therefore, although the overall experience of the entire contributor community may be important, the top contributor's platform and community-specific experience may be particularly critical. A greater depth of experience, in terms of previous experience editing on Wikipedia, will be positively related to the quality of peer production in the community. This rationale leads to a fourth hypothesis.

*H4a.* The top contributor's depth of experience will be positively related to the quality of peer-produced information.

Nevertheless, the total number of articles to which the contributor has contributed may offset the value of this experience for the collaborative process. If the top contributor typically makes a few contributions to a large number of articles, it is unlikely that this contributor will have substantively engaged in other collaborative processes and learned how to be more effective as a result of this experience. It may mean that the contributor's involvement is shallow or focused around a particular issue or role and will be unable to offer the type of leadership as collaborators with greater depth of experience. For instance, some contributors to the VT massacre article indicated that they worked only on issues related to gun control and others simply scanned articles for typos. If these editors are the top contributors to content, it is unlikely that they will bring the type of leadership and guidance that individuals with deeper experience might. The breadth of contributions also may mean that these contributors do not have a great deal of commitment to develop content, since they are frequently moving from community to community rather than investing in the efforts of a small few.

*H4b.* The top contributor's breadth of experience will be negatively related to the quality of peer-produced information.

#### 4. QUANTITATIVE ANALYSIS OF FEATURED ARTICLE NOMINATIONS

Phase 2 of the research sought to test whether the hypotheses developed in the study of the VT massacre article would hold across other Wikipedia articles. Data were collected on the collaborative processes behind 188 Wikipedia articles that had been nominated for featured article status during the time period studied during Phase 1. Although Wikipedia awarded feature article status to the VT massacre article, featured articles do not necessarily represent more popular or higher-profile articles. Featured article status simply recognizes that the information quality of the article exemplifies the best of what Wikipedia hopes to accomplish. Other collaborative environments may define successful outcomes differently, and these settings should be evaluated on those explicit standards. Nevertheless, the Wikipedia standards for information quality, that is "complete, accurate, and well-written," likely apply to many collaborative settings.

Several types of data were collected on these articles. First, the research team used an automated data collection tool to compile statistics on each article's contributor history. This tool provided the number and type of contributions, contributors' identities and activities, the date the article was created, and a breakdown of contributor activity by

month. Data from both the article and the discussion pages dedicated to the article were combined to provide the broadest picture of collaboration relating to each article. Second, the author used another automated tool to capture the contribution history of the top contributor of each article.

#### 4.1. Measures

From these data, I constructed several variables of interest for analysis. Appendix A shows descriptive statistics for these variables, and Appendix B shows the correlation matrix. The dependent variable is a binary indicator of whether a nominated article was promoted to featured article status (1) or whether it was rejected (0).

I used several measures to capture volume of collaborative activity. Total contributors was captured by averaging the number of contributors between each article page and discussion page. Average contributions per contributor was calculated by adding the total number of contributions to each article and the discussion pages divided by the total contributors variable constructed before to avoid multicollinearity problems. The total amount of shaping or packaging that occurred in the article was captured by minor contributions, calculated as a percentage by dividing the number of minor contributions by the total contributions on both the article and discussion pages. When contributors make edits, they can select a checkbox to indicate that their edits are minor. Although these data were self-reported, a random audit confirmed that they were minor.

Anonymous contributors was also calculated as a percentage by dividing the number of distinct IP addresses that made contributions by the total contributors variable. A single contributor may use different IP addresses, so that researchers and Wikipedia contributors cannot know that the contributions are coming from the same person, so anonymous IP address were treated as representing a separate contributor. Anonymous contributors was chosen (rather than volume of anonymous contributions), because it is often possible to identify that a string of activity with a particular IP address represents a single contributor. These contributions typically involve a single discussion or edit, so that other contributors can treat the contributions as coming from one contributor.

Top contributor experience was captured by examining the top contributor's career in terms of two variables. Experience depth was captured through key players' number of contributions on all Wikipedia pages since first joining Wikipedia, which indicated overall commitment to Wikipedia. Experience breadth was captured through contributors' number of unique pages edited since first joining Wikipedia, which indicated project ranges. These variables employ data from from the founding of Wikipedia through April 2007.

We controlled for various article characteristics that may influence the perceptions of information quality. Readers may perceive that longer articles are of higher quality (i.e., more information is better), so our models controlled for word count, operationalized as the number of words in the article. More extensively referenced articles may lead to perceptions that content is more authoritative without regard to content quality, so our models controlled for reference count, operationalized as the number of unique citations in the article. Multimedia images may increase the visual appeal of an article, so our models controlled for picture count, operationalized as the total number of multimedia images, maps, and other graphics. Article age may be an important factor, as more time allows collaboration to move at a more deliberative pace, operationalized as the number of days from the founding to nomination for featured article. It also may be important to know what percentage of total contributions are direct contributions compared to coordinating activity [Kittur and Kraut 2008], so the models control for percentage of article collaboration to determine what portion of the total edits happen on the article versus the talk page.

Table III. Results

	<i>Model 1</i>			<i>Model 2</i>		
	B	S.E.	p-value	B	S.E.	p-value
Constant	0.376*	0.149	0.012	0.568***	0.17	0.001
Word Count	-0.587*	0.241	0.015	-0.725**	0.279	0.009
Picture Count	0.096	0.198	0.628	0.001	0.218	0.996
Reference Count	0.523*	0.215	0.015	0.874***	0.267	0.001
Article Age	0.023	0.163	0.888	0.126	0.217	0.563
Pct. Article Collaboration	-0.157	0.149	0.291	-0.183	0.184	0.322
Total Contributors				0.279	0.254	0.271
Average Contributions				-0.017	0.211	0.934
Minor Contributions				0.416*	0.211	0.049
Anonymous Contributors				-0.467*	0.232	0.044
Top Contributor Experience (Depth)				0.931*	0.374	0.013
Top Contributor Experience (Breadth)				-1.005**	0.361	0.005
Nagelkerke R <sup>2</sup>		.064			.217	

\* =  $p < .05$ , \*\* =  $p < .01$ , \*\*\* =  $p < .001$ .

#### 4.2. Data Analysis and Results

The data were analyzed using the binary logistic regression analysis in SPSS. The data did not exhibit characteristics of multicollinearity. The Hosmer and Lemeshow test showed that the models adequately fit the data. Table III shows the results of the data analysis. Model 1 contains only the control variables, and Model 2 includes the independent variables of interest. Models were tested where each independent variable of interest was entered individually, with no effects on the results presented here.

Hypothesis 1, which was not supported, stated that the activity of the contributor community would be positively related to the quality of peer-produced information. This finding is surprising, considering that previous research has identified volume of contributor activity to be the one feature of collaboration that is related to the quality of peer-produced information on Wikipedia [Ball 2007; Kittur et al. 2007; Wilkinson and Huberman 2007]. Previous research, however, has compared featured articles with all other articles on Wikipedia. A great many articles on Wikipedia, however, may be simply user pages that are dedicated to a given user or may be incomplete proto-articles comprising only a few sentences (i.e., article stubs) created accidentally when users are prompted to create entries if their search fails to retrieve a reference. The volume of contributor activity may be an important first-cut in determining a base level of the quality of peer-produced information, but it may be unimportant for determining quality beyond this most basic level.

Hypothesis 2, which was supported, stated that the amount of effort spent shaping the article through minor contributions would be positively related to the quality of peer-produced information. This finding suggests that the amount of energy contributors spend on content shaping is positively related to the quality of peer-produced information.

Hypothesis 3 stated that the number of anonymous contributors on a given article would be negatively related to the quality of peer-produced information. Hypothesis 3 was supported. Phase 1 research revealed that anonymous contributors were associated with higher levels of vandalism, requiring the contributor community to spend time and energy to assess their veracity or combat their damage.

Hypothesis 4 stated that the top contributor's experience is likely to be related (depth positively related, breadth negatively related) to the quality of peer-produced information. These hypotheses are supported. If the most prolific contributor has deeper experience in the Wikipedia environment, it is more likely that the community will develop the highest-quality content. If the top contributor, however, shows broad involvement

indicative of superficial or niche involvement, it is less likely that the community will develop the highest-quality content.

## 5. DISCUSSION AND CONCLUSION

Until now, little has been known about effective collaborative processes in wiki environments. This study uses qualitative data from an in-depth case study of the Wikipedia article on the VT massacre to motivate the quantitative analysis of the collaborative processes of 188 Wikipedia articles. Providing a rich picture of wiki collaboration, these qualitative data are then used to develop testable hypotheses that are tested against the quantitative data from 188 articles nominated for featured article status during the time period covered in Phase 1. This analysis confirms certain features that are positively associated with the quality of peer-produced information (content shaping, depth of top contributor experience), some that are negatively associated (anonymous contributors, breadth of top contributor experience), and one that is, surprisingly, found to have no relationship with the quality of peer-produced information (total volume of contributions and contributors).

This article has a number of implications for research and practice. The findings, combined with previous research, suggest that a critical mass of contributors and participants may be important to reach a base level of collaborative success; but additional contributors may do little to improve collaborative success once the critical mass is achieved. Thus, it is important to attract and retain enough contributors, but organizations should turn their attention to other collaborative features once they have established a sufficiently large base of contributors. Managers and leaders of online communities would be wise to avoid a “more is better” mentality. While attracting a critical mass of contributors may be the initial and most pressing task faced in peer-production environments, it is by no means the only one. Further research might investigate whether and where this critical mass threshold exists.

Second, this study suggests that the composition of the contributor community is critical for successful collaborative efforts. Anonymous contributors introduce a number of difficulties into the collaborative process, primarily because the community must evaluate the quality of their contributions. Being able to identify the contributor (along with contributor reputation or history) appears to be critical in evaluating contributions. Although anonymity may provide value in other collaborative environments (i.e., brainstorming intended to generate novel ideas), it negatively affects the wiki-based collaboration environment studied here. Managers, therefore, should approach anonymity in peer-production communities cautiously. This research suggests that even if anonymity does not directly and detrimentally inhibit collaborative outcomes, it may indirectly compromise the work of the community.

Third, although substantive contributions are important for wiki-based collaboration, we should not overlook the importance of minor contributions targeted at shaping or packaging knowledge. Although this research does not explain the contributor community’s approaches to content shaping (e.g., do some contributors focus more on copyediting than contributing? Is packaging always the core community’s function?), it does find that attention to these stylistic details is an important factor in effective wiki-enabled collaboration. This finding is theoretically significant in that some researchers have identified “shaping” behaviors as the key theoretical distinction that separates wikis from other types of collaborative tools [Kane et al. 2009b; Yates et al. 2010]. Managers and community leaders may be tempted to focus on generating information only, but they should pay equal if not greater attention to how they can best tailor this information for the intended consumers.

Fourth, top contributors’ experience is also important because these top contributors are likely to be the community leaders [Lave and Wenger 1991; Preece and

Schneiderman 2009]. When these leaders have rich experience outside the immediate collaborative environment, they can bring this experience to bear on the collaboration at hand and improving the quality of collaboration [Kane 2009]. On the other hand, when these leaders have broad experience outside the community, it appears to detract from the quality of the collaboration. It may be that these contributors are dilettantes, not committed to the development of quality content, or they may have a particular set of nontopical interests (e.g. gun control, copyediting) that do not translate well into the leadership role they find themselves in. Managers and community leaders may find it beneficial to intentionally seed the community with deeply experienced leaders or provide support and training to emergent leaders to facilitate information creation in similar settings. Further research might explore the nature of effective leadership in online production communities.

Some limitations must be considered when assessing the contribution of this work. First, the topic may be an important variable associated with information quality, but the research method did not allow for control of possible topical differences. Different collaborative processes may be associated with certain topics. For instance, medical articles have collaborative standards that differ from those established for typical Wikipedia articles [Kane 2009], which may lead to different collaborative patterns associated with information quality. Second, although research has suggested that findings regarding collaboration on Wikipedia generalize to collaboration in other social media settings [Kittur and Kraut 2010], the generalization is certainly imperfect; any findings should be generalized carefully to other settings. Further research is needed to establish whether these findings generalize to corporate settings. Third, the particular Wikipedia articles chosen may or may not reflect typical collaborations on Wikipedia. Many peer-produced collaboration environments have problems simply creating enough collaboration activity to generate information. Because only articles nominated for featured article status were sampled, all the sample articles had overcome this base problem, by definition. Nevertheless, since most companies are not seeking to just create more information with peer-production communities, but high-quality information with considerable economic value, the distinction between complete articles and the best ones created by peer-production communities is one worth studying.

In conclusion, this study examines effective peer-production processes on a wiki platform. Companies and nonprofit organizations are increasingly using wiki platforms to enable employees, customers, or volunteer contributors to collaborate. Some wiki-based collaboration efforts have yielded remarkable successes; others have been colossal failures. Thus, it is important for companies to understand what factors of wiki-based collaborations are associated with the creation of valuable and high-quality information.

## APPENDIX

Appendix A. Descriptive Statistics

	N	Min.	Max.	Mean	Std. Deviation
Pass Nomination	195	0	1	.59	.493
Word Count	195	135	11078	4527.32	2015.166
Picture Count	195	0	42	8.64	7.182
Reference Count	195	4	206	66.13	36.034
Article Age	195	4	2257	1183	644.836
Pct. Article Collaboration	195	.46	1.00	.871	0.086
Total Contributors	195	8	2317	278.02	405.81717
Average Contributions	195	3.70	44.30	8.7839	6.45707
Anonymous Contributors	195	.00	.67	.3704	.16661
Minor Contributions	195	.06	.74	.2674	.11716
Top Contributor Breadth	188	14	25878	3156.60	3556.001
Top Contributor Depth	188	134	45000	9967.21	8904.667

Appendix B. Correlation Matrix

	1	2	3	4	5	6	7	8	9	10	11	12
1 Pass Nomination	1											
2 Word Count	-0.08	1										
3 Picture Count	0.008	.621**	1									
4 Reference Count	0.085	.660**	.513**	1								
5 Article Age	-0.015	.381**	.333**	.285**	1							
6 Pct. Article Collaboration	0.088	0.023	0.077	0.012	-0.062	1						
7 Total Contributors	-0.022	.430**	.390**	.454**	.490**	-.172*	1					
8 Average Contributions	0.012	.409**	.389**	.468**	.439**	-.330**	.918**	1				
9 Anonymous Contributors	-.165*	.289**	.185**	.342**	.485**	0.006	.533**	.453**	1			
10 Minor Contributions	.182*	-0.051	-0.051	-0.129	-0.083	.157*	-.203**	-.213**	-.408**	1		
11 Top Contributor Breadth	-0.031	0.035	0.03	0.057	0.002	0.053	0.108	0.046	-0.069	.196**	1	
12 Top Contributor Depth	0.095	0.066	0.057	0.089	0.007	0.071	0.081	0.061	-0.102	.218**	.820**	1

\*p &lt; .05, \*\*p &lt; .01.

## REFERENCES

- ALAVI, M. AND LEIDNER, D. E. 2001a. Research commentary: Technology-Mediated learning—A call for greater depth and breadth of research. *Inf. Syst. Res.* 12, 1–10.
- ALAVI, M. AND LEIDNER, D. E. 2001b. Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *Mis Quart.* 25, 107–136.
- BALL, P. 2007. The More, the Wikier. *Nature*.
- BOCK, G. W., ZMUD, R. W., KIM, Y. G., AND LEE, J. N. 2005. Behavioral intention formation in knowledge sharing: Examining the roles of extrinsic motivators, social-psychological forces, and organizational climate. *Mis Quart.* 29, 87–111.
- BUTLER, B. S., JOYCE, E., AND PIKE, J. 2008. Don't look now, but we've created a bureaucracy: The nature and roles of policies and rules in wikipedia. In *Proceedings of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems*. 1101–1110.
- CONNOLLY, T., JESSUP, M. L., AND VALACICH, J. S. 1990. Effects of anonymity and evaluative tone on idea generation in computer mediated groups. *Manag. Sci.* 36, 689–703.
- DEE, J. 2007. All the news that's fit to print out. *New York Times*.
- GRIFFITH, T. L., SAWYER, J. E., AND NEALE, M. A. 2003. Virtualness and knowledge in teams: Managing the love triangle of organizations, individuals, and information technology. *Mis Quart.* 27, 265–287.
- JESSUP, L. M., CONNOLLY, T., AND GALEGHER, J. 1990. The effects of anonymity on Gdss group-process with an idea-generating task. *Mis Quart.* 14, 313–321.
- KANE, G. C. 2009. It's a network, not an encyclopedia: A social network perspective on Wikipedia collaboration. In *Proceedings of the Academy of Management Annual Meeting*, G. T. Solomon, Ed.
- KANE, G. C. AND FICHMAN, R. G. 2009. The shoemakers children: Using wikis for IS research, teaching, and publication. *Mis Quart.* 33, 1–22.
- KANE, G. C., FICHMAN, R. G., GALLAUGHER, J., AND GLASER, J. 2009a. Community relations 2.0: With the rise of real-time social media, the rules about community outreach have changed. *Harvard Bus. Rev.* 87, 45–50.
- KANE, G. C., MAJCHRZAK, A., JOHNSON, J., AND CHEN, G. L. 2009b. A lifecycle model of perspective making and perspective taking in fluid online collectives. In *Proceedings of the International Conference on Information Systems*.
- KANKANHALLI, A., TAN, B. C. Y., AND WEI, K. K. 2005. Contributing knowledge to electronic knowledge repositories: An empirical investigation. *Mis Quart.* 29, 113–143.
- KITTUR, A., CHI, E., PENDLETON, B. A., SUH, B., AND MYTKOWICZ, T. 2007. Who writes Wikipedia? How centralization scales (and doesn't) in social collaborative systems. Working Paper.
- KITTUR, A. AND KRAUT, R. E. 2008. Harnessing the wisdom of crowds in Wikipedia: Quality through coordination. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work (CSCW'08)*. ACM Press, New York.
- KITTUR, A., LEE, B., AND KRAUT, R. E. 2009. Coordination in collective intelligence: The role of team structure and task interdependence. In *Proceedings of the Conference on Computer Human Interaction (CHI'09)*.
- KITTUR, A. AND KRAUT, R. E. 2010. Beyond Wikipedia: Coordination and conflict in online production groups. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*. ACM, 215–224.
- LAVE, J. AND WENGER, E. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, Cambridge, UK.
- LI, C. AND BERNOFF, J. 2008. *Groundswell: Winning in a World Transformed by Social Technologies*. Harvard Business Press, Boston, MA.
- MARKUS, M. L. 2001. Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success. *J. Manag. Inf. Syst.* 18, 57–93.
- MILES, M. B. AND HUBERMAN, A. M. 1994. *Qualitative Data Analysis: An Expanded Sourcebook*. Sage Publications, Thousand Oaks, CA.
- ORTEGA, F., GONZALEZ-BARAHONA, J. M., AND ROBLES, G. 2008. On the inequality of contributions to Wikipedia. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*.
- PREECE, J. AND SCHNEIDERMAN, B. 2009. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Trans. Hum.-Comput. Interact.* 1, 13–32.
- RAYMOND, E. S. 1999. *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. O'Reilly, Cambridge, MA.
- SARNER, A. 2008. The business impact of social computing on marketing and 'generation virtual'. <http://www.gartner.com/it/page.jsp?id=770914>

- SHORT, J., WILLIAMS, E., AND CHRISTIE, B. 1976. *The Social Psychology of Telecommunications*. John Wiley, New York.
- SIA, C. L., TAN, B. C. Y., AND WEI, K. K. 2002. Group polarization and computer-mediated communication: Effects of communication cues, social presence, and anonymity. *Inf. Syst. Res.* 13, 70–90.
- SIEGEL, J., DUBROVSKY, V., KIESLER, S. B., AND MCGUIRE, T. W. 1986. Group processes in computer-mediated communication. *Ograniz. Behav. Hum. Decis. Process.* 37, 157–187.
- STEIN, E. W. AND ZWASS, V. 1995. Actualizing organizational memory with information-systems. *Inf. Syst. Res.* 6, 85–117.
- TAPSCOTT, D. AND WILLIAMS, A. D. 2006. *Wikinomics: How Mass Collaboration Changes Everything*. Portfolio Hardcover.
- VALACICH, J. S., DENNIS, A. R., AND NUNAMAKER, J. F. 1992. Group size and anonymity effects on computer-mediated idea generation. *Small Group Res.* 25, 83–104.
- WAGNER, C. AND MAJCHRZAK, A. 2006. Enabling customer-centricity using wikis and the wiki way. *J. Manag. Inf. Syst.* 23, 17–43.
- WILKINSON, D. M. AND HUBERMAN, B. A. 2007. Assessing the value of cooperation in Wikipedia. Tech. rep., HP Labs, Palo Alto, CA.
- YATES, D., WAQNER, C., AND MAJCHRZAK, A. 2010. Factors affecting shapers of organizational wikis. *J. Amer. Soc. Inform. Sci. Technol.* 61, 543–554.
- YIN, R. K. 2008. *Case Study Research: Design and Methods* 4th Ed. Sage Publications, Thousand Oaks, CA.

Received April 2010; revised December 2010; accepted December 2010