WHEN DOES CROWD SIZE MATTER?

THE INFLUENCE OF DIVERSITY AND EXPERIENCE ON THE EFFECTS OF CROWD SIZE

Lionel P. Robert Jr.[1,2]
School of Information
University of Michigan
4381 North Quad
105 S. State St.
Ann Arbor, MI 48109-1285
Phone: 734-764-5296
Fax: 734-615-3587
lprobert@umich.edu

Daniel M. Romero [2]
School of Information
University of Michigan
4337 North Quad
105 S. State St.
Ann Arbor, MI 48109-1285
Phone: 734-764-4096
Fax: 734-615-3587
drom@umich.edu

[1] Corresponding Author
[2] Lionel P. Robert Jr. and Daniel M. Romero contributed equally to this work.

WHEN DOES CROWD SIZE MATTER?

THE INFLUENCE OF DIVERSITY AND EXPERIENCE ON THE EFFECTS OF CROWD SIZE

Abstract

One advantage of crowds over traditional teams is that crowds enable the assembling of a large number of individuals to address problems. The literature is unclear, however, about when crowd size leads to better outcomes. To better understand the effects of crowd size we conducted a study on the retention and performance of 4,317 articles in the WikiProject Film community. Results indicate that crowd composition, specifically diversity and experience, is vital to understanding when size leads to better retention and performance. Crowd size was positively related to retention and performance when crowds were high in diversity and experience. Retention was important to determining when crowd size led to better performance. Crowd size was positively related to performance when retention was low. Our results suggest that crowds benefit from their size when they are diverse, experienced, and have low retention rates.

Introduction

Crowds are being deployed to tackle complex problems and issues that were once the domain of traditional organizational teams (Chen, Ren, & Riedl, 2010; Yu & Nickerson, 2011). Yet, unlike their organizational counterparts, we know much less about what facilitates better retention and performance in crowds (Qin, Salter-Townshend, & Cunningham, 2014). As a result, our knowledge of what leads to effective crowd work has not kept pace with the deployment of crowds in our global economy (Ransbotham & Kane, 2011; Xiao, 2014). As crowds become increasingly vital to how work gets done, there is a need to better comprehend what conditions facilitate better retention and performance (Arazy, Yeo, & Nov, 2013).

One advantage of crowds over traditional teams is that they enable the assembling of large numbers of individuals. Yet the literature is unclear about when crowd size might actually lead to better outcomes. There are three views on the relationship between size and both retention and performance in the literature on traditional teams. One view is that size should increase retention and performance. As size increases so should work output and resources available to attract more members (Cummings, Kiesler, Zadeh, & Balakrishnan, 2013; Hausknecht, Trevor, Howard, 2009). Another view is that size should be negatively related to retention and performance. This view asserts that as size increases coordination cost and conflict increase and individual effort decreases, all of which should lower retention and performance (Alnuaimi, Robert, & Maruping, 2010; Hausknecht et al., 2009; Newell et al., 2008; Srinivasan et al., 2012, Romero et al. 2015). A third view on size highlighted by Oliver and Marwell (1988) is that the effects of size are relative to a group's composition.

This paper builds on this third view and examines the role of crowd composition on the relationship between crowd size and both retention and performance. This paper extends the literature on crowds in three ways. One, we examine the how crowd composition, in terms of diversity and experience, impacts the relationship between size and performance. Two, we examine the impact of diversity and experience on the relationship between crowd size and retention. Retention, the ability to retain participants in a crowd, has

3

been viewed as an important outcome in crowd work (Halfaker, Geiger, Morgan & Riedl 2013; Ransbotham & Kane, 2011). However, research has not examined how size, diversity and experience can combine to undermine or facilitate better retention in crowds. Three, we investigate the relationship between retention and performance. Although retention has been touted as a key component to successful crowd work, few studies have directly examined when retention is good or bad for crowd performance (Ransbotham & Kane, 2011). To examine these relationships, we conducted a study on the quality of 4,317 articles in the WikiProject Film community. Our findings provide new insights into the relationships among crowd size, diversity, experience, retention, and performance.

In the following section, we review the literature on the relationship between crowd size and performance in Wikipedia crowds. Then we discuss the current literature on retention in Wikipedia crowds and the need for more research on the effects of crowd size on retention. Next, we present our research model, where we discuss the theoretical linkages among crowd size, crowd composition, retention and performance. Then, we present the methods and results sections. Finally, we discuss the implications of our findings.

Background and Theoretical Model

Crowd size and the performance of crowds

The research on crowd size and crowd performance in Wikipedia crowds can be divided into three categories: the direct effect approach, the indirect effect approach, and the contingency approach. The direct effect approach includes studies that examine the relationship between crowd size and performance in crowds without any mediators or moderators. Some of these studies employ crowd size as one of the theoretical constructs of interest while others include it as only a control variable. The results have been inconsistent. For example, several studies have included crowd size as a control variable predicting performance, and these studies found that crowd size is non-significant (Arazy & Nov, 2010; Arazy, Nov,

Patterson, & Yeo, 2011; Kane, 2011). However, several other studies found that size is positively related to crowd performance (Arazy et al., 2013; Carillo & Okoli, 2011; Choi, Woo, & Han, 2013; Wilkinson & Huberman, 2007). The indirect approach involves studies that examine how crowd size impacts crowd performance through a mediating variable. One study employing the indirect approach found that crowd size decreased crowd performance, in part by increasing diversity (Arazy, Morgan, & Patterson, 2006).

The contingency approach examines how moderators influence the relationship between crowd size and performance. Several studies have employed the contingency approach to understand the relationship between crowd size and performance. Two of these studies examined the influence of inequality of workload among editors on the relationship between crowd size and performance (Kittur & Kraut, 2008; Kittur, Lee, & Kraut, 2009). Low levels of inequality in article workload mean that work is evenly distributed across all editors, while high levels of inequality mean that work is unevenly distributed and that most work is done by a relatively few editors. These studies argue that inequality in article workload represents better or more coordination, which should become increasingly important as crowd size increases. They have found that inequality in article workload either decreased the negative relationship between size and performance (Kittur & Kraut, 2008) or strengthened the positive relationship between crowd size and performance (Kittur et al., 2009). Another study found that discussion among crowd members moderated the relationship between crowd size and performance. Specifically, the positive relationship between crowd size and performance became stronger when discussion among crowd members increased (Choi et al., 2013). In all, the literature on crowd size suggests that the relationship between size and performance is likely to be dependent on other variables.

Crowd size and retention in crowds

Retention remains an important topic to Wikipedia scholars (Mesgari, Okoli, Mehdi, Nielsen and Lanamäki, 2015). In the near term, the importance of retention is demonstrated by the strong link between

retention and performance (Ransbotham & Kane, 2011). In the long term, the ability to attract and maintain a voluntary workforce is critical to the success of Wikipedia's long-term sustainability (Halfaker et al., 2013. Yet, scholars have noted that there has been a decrease in the number of active participants in the Wikipedia community (Lam & Riedl, 2011). This decrease has been attributed to Wikipedia's inability to retain participants in its community (Halfaker et al., 2013. The importance of retention is expected to increase as the topic of recruitment and inclusion of under-represented minorities becomes central to the broader Wikipedia community (WMDW, 2014).

Much of the research on retention has focused on policies and procedures that can create conditions that facilitate retention (see Halfaker et al., 2013, for a review). However, much less research has been directed at understanding how crowd characteristics might influence retention. This is in stark contrast to previous studies examining retention in other online groups, where group characteristics like size have shown to be particularly important in understanding retention (Butler, 2001). Similarly, we know very little about how retention might impact the performance of crowds. For example, in a recent review of Wikipedia research, Mesgari et al. (2015) only identified one study that examined the relationship between retention and crowd performance. Surprisingly, the study showed that low levels of retention could actually be good for performance under certain circumstances (Ransbotham & Kane, 2011). In all, more research is needed to better understand the relationship between crowd size and retention, and retention and crowd performance.

Theoretical model

Our model asserts that the relationship between crowd size and both retention and performance is dependent largely on crowd composition (in this study defined as diversity and experience). Our model further asserts that the relationship between crowd size and performance is also largely dependent on retention (see Figure 1). In developing our hypotheses, we draw from the literature on both crowds and

traditional organizational teams. However, our hypotheses are specific to the Wikipedia crowds examined in our study.
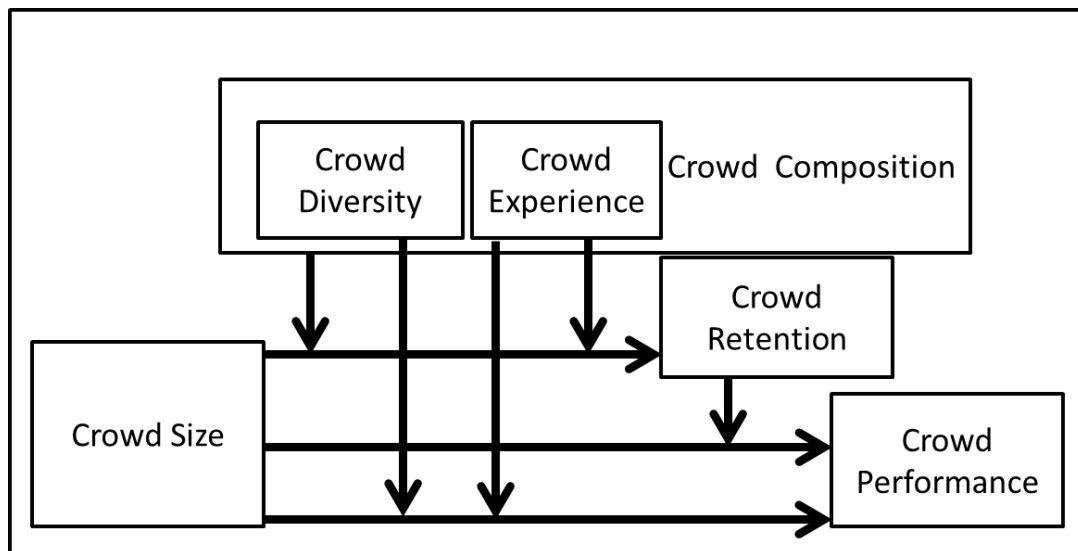

Figure 1. Research Model

We examine three types of diversity found in the literature examining Wikipedia crowds (Arazy & Nov, 2010; Chen et al., 2010; Kittur & Kraut, 2008). The first two can be considered a type of disparity. Disparity represents the inequality in power, roles, and resources (Harrison & Klein, 2007). The first disparity construct is inequality in article workload when considering the work performed within each article. We label this construct as *local workload diversity* (LWD). This construct originates from Kittur and Kraut (2008) and was originally used as a measure of coordination. The second is disparity in terms of the inequality in workload when considering the work performed in all of Wikipedia. We label this construct as *global workload diversity* (GWD). GWD was inspired from Arazy and Nov (2010) and represents the inequality in broader knowledge of the Wikipedia community. LWD and GWD both represent a measure of disparity in workload. However, because one measures disparity in *within* article workload and the other one measures disparity in workload over all of Wikipedia, they could have very different effects. Indeed, users with high workload within an article could have little workload across Wikipedia if they only contribute to a single or very few articles, and hence, articles with high LWD could have low GWD and vice versa.

Users with high levels of contributions within an article are most likely experts in the subject of the article and users with high levels of contribution in Wikipedia are likely to be experts in Wikipedia rules and conventions. Therefore, it is important to distinguish and separately measure LWD and GWD. The third diversity construct can be considered a type of variety. It represents the differences in information, knowledge, and backgrounds among team members (Harrison & Klein, 2007). We operationalized it in terms of the number of topic areas editors had previously worked on together (Chen et al., 2010). This is labeled as *interest diversity*. Taken together, we examine crowd diversity as LWD and GWD along with interest diversity.

Crowd size and local workload diversity (LWD)

LWD should moderate the relationship between crowd size and performance. As crowd size increases, it becomes more difficult to coordinate work among individuals (Cummings et al., 2013); yet, at the same time this coordination becomes more important (Kittur & Kraut, 2008). In fact, the inability to coordinate is often cited as one of the major reasons that increases in group size are often associated with decreases in performance (Thomas & Fink, 1963; Weber, Camerer, Rottenstreich, & Knez, 2001). However, in crowds with high LWD, only a relatively small subset of individuals has to coordinate their work (Arazy & Nov, 2010). Hence, in crowds with high LWD, coordination efforts do not have to increase at the same rate as crowd size (Arazy & Nov, 2010). This allows crowds to benefit from increases in size without the drawbacks associated with the inability to coordinate work (Kittur & Kraut, 2008).

H1a) Crowd size is positively related to crowd performance when LWD increases.

Crowd size should be positively related to retention when LWD is high, for two reasons. One, increases in crowd size can create a chaotic and unstructured work environment that newcomers find difficult to assimilate into because of coordination problems (Ransbotham & Kane, 2011). This often leads

to re-edits — the actions taken by one editor to overwrite the edits of another editor — and the so-called *edit wars* that often cause decreases in crowd retention (Halfaker, Kittur, & Riedl, 2011; Suh, Convertino, Chi, & Pirolli, 2009). Because crowds high in LWD have considerable variance in the amount of work done across participants, they typically have the property that a few individuals produce more of the work (Kittur et al., 2009). This core set of editors can provide structure to an otherwise unstructured environment, making edit wars less likely or more manageable.

Two, as crowd size increases it becomes difficult for newcomers to understand the implicit norms and unwritten rules of the crowd, such as when and where to contribute (Krieger, Stark, & Klemmer, 2009), which decreases retention (Halfaker, Geiger, Morgan & Riedl, 2013). The core group of editors present in crowds high in LWD becomes a strong reference point. As such, when newcomers need to understand the implicit norms and routines of the crowd they have a core set of editors to both model themselves after and go to for questions. Both the ordered work environment and the ability to have a strong reference group should help crowds leverage their size to retain more members.

H1b) Crowd size is positively related to crowd retention when LWD increases.

Crowd size and global workload diversity (GWD)

GWD should also help crowds take advantage of their increases in size. First, GWD can represent the ability of crowds to manage work across articles. Wikipedia articles are not standalone entities but instead have to fit within the broader Wikipedia community (Kittur & Kraut, 2010). As crowd size increases, it can become a challenge to ensure consistency across articles. Editors with experience in the broader Wikipedia community are more likely to be in a position to help crowds communicate and coordinate across articles to ensure consistency (Kittur & Kraut, 2008).

Two, GWD may also enhance the capability of crowds to manage conflict and resolve deadlocks. Conflict can both result from and lead to work stoppages or other behavior detrimental to crowd

performance (Arazy et al., 2013). As crowd size increases such events are likely to increase in frequency (Kittur & Kraut, 2008), but editors with experience in the broader Wikipedia community are likely to know how to handle problems related to conflict or other unproductive behaviors (Arazy & Nov, 2010). This assertion aligns with other research on Wikipedia crowds that found that GWD leads to better crowd performance (Arazy & Nov, 2010). This is also consistent with findings that diversity in unequal power distributions facilitates hierarchical structures, which can lead to more effective decision-making and ultimately performance (Daniel, Agarwal & Stewart, 2013; Groysberg, Polzer, & Elfenbein, 2011; Overbeck, Correll, & Park, 2005). In sum, as the process losses associated with size increase, increases in GWD should help crowd size lead to better crowd performance.

H2a) Crowd size is positively related to crowd performance when GWD increases.

A similar relationship should exist regarding crowd size, GWD and retention. As crowd size increases so does the likelihood that conflict and other unproductive activities will occur (Kittur & Kraut, 2010). Having a group of experienced editors who can intervene to resolve these issues and provide guidance to the crowd is much more inductive to retaining participants. For example, research on Wikipedia crowds has shown that retention increases when crowds can resolve disputes (Halfaker et al., 2011). When increases in crowd size are accompanied by increases in GWD we expect crowd size to be associated with better retention.

H2b) Crowd size is positively related to crowd retention when GWD increases.

Crowd size and interest diversity

We also assert that diversity in members' interests should determine whether increased crowd size leads to better performance. The assertion that crowd size is positively related to crowd performance is often built on the assumption that more individuals can bring additional skills and knowledge to the group

(Van Knippenberg & Schippers, 2007; Windeler, Maruping, Robert & Riemenschneider, 2015). When this is true, increases in size lead to increases in the capabilities of the crowd, which in turn should lead to better crowd performance. However, this is only true if each new member brings additional unique knowledge or skills to the crowd (Page, 2007). Diversity in members' interests can reflect the differences in knowledge and skills among members (Chen et al., 2010). Therefore, increases in crowd size should be positively related to crowd performance when interest diversity is high.

On the other hand, when crowds are low in interest diversity, increases in crowd size are less likely to translate to better crowd performance. When crowds low in interest diversity increase in size they are adding new members with the same knowledge or skill set (Van Knippenberg & Schippers, 2007). These crowds are likely to encounter social loafing as size increases (Alnuaimi, Robert, & Maruping, 2009; Srinivasan et al., 2012). Social loafing, or the reduction of individual effort within groups, has been shown to increase as group size increases (Alnuaimi et al., 2010; Srinivasan, Maruping, & Robert, 2010). However, social loafing is less likely to occur in groups with diverse information and knowledge (Valacich, Wheeler, Mennecke, & Wachter, 1995).

H3a) Crowd size is positively related to crowd performance as interest diversity increases.

Along with enabling crowd size to lead to better performance, interest diversity should also help crowd size lead to more retention. As crowds increase they can become daunting for newcomers. Increases in traditional team size have been associated with less cohesion and more conflict (Amason & Sapienza, 1997). Crowds with more interest diversity are likely to avoid these problems normally associated with increases in size. Newcomers who join groups typically look for others who have similar interests (Gotsis & Kortezi, 2015; Newell, Robert, Riemenschneider, & Maruping, 2009; Robert 2013). These newcomers are much more likely to find others with similar interests in crowds high in interest diversity. Additionally, groups high in interest diversity are often more tolerant and accepting of others (Cox

11

& Blake, 1991; Gonzalez & Denisi, 2009). The same could be true for crowds. Therefore, we might expect highly diverse crowds to be more open to newcomers. Taken together, both the ability to find like others and the more tolerant climate in crowds with more interest diversity should allow increases in crowd size to convert to more retention.

H3b) Crowd size is positively related to crowd retention as interest diversity increases.

Crowd size and experience

More experienced editors should be better able to personally deal with issues associated with increases in crowd size and help others deal with these issues also. Research in organizations has consistently linked employee experience to higher levels of knowledge and skills (Dokko, Wilk, & Rothbard, 2009; Tesluk & Jacobs, 1998; Wright & Bonett, 2002) and more willingness to help others perform their job well (Ng & Feldman, 2010). More experienced editors should also be less susceptible to effort reduction as crowd size increases. Experienced employees are often the most motivated employees (Bretz & Judge, 1994; Tesluk & Jacobs, 1998; Wagner, Ferris, Fandt, & Wayne, 1987). This is particularly important because decreases in individual effort, which can occur as size increases, have been accredited to decreases in motivation (Alnuaimi et al., 2010). This means that crowds with more experienced editors are less susceptible to decreases in individual effort as size increases. In sum, when crowds have editors who have more ability, are willing to help others, and are more motivated, increases in size are more likely to be positively related to crowd performance.

H4a) Crowd size is positively related to crowd performance as crowd experience increases.

Crowd size should also be positively related to retention when crowds have more experience. Experienced members are both less likely to leave and more likely to help retain new members. Organizational scholars argue that experienced employees are less sensitive to the changing work

conditions and the stress that often leads to turnover (Johnston, Parasuraman, & Futrell, 1989). One reason this occurs is that employees with more experience are likely to have developed coping skills to handle stressful situations (Armstrong-Stassen, 1994; Long, 1993; Wright & Bonett, 1993), and stress does not necessary lead to turnover when individuals can find ways to cope with it (Wright & Bonett, 1993). Another reason is that experienced employees are more committed than inexperienced ones (Joseph, Kok-Yee, Koh, & Soon, 2007). When difficulties occur, more committed employees are less likely to leave (Ahuja, Chudoba, Kacmar, McKnight, & George, 2007). Both better coping skills and higher levels of commitment should translate into higher retention when crowd size increases.

H4b) Crowd size is positively related to crowd retention as crowd experience increases.

Crowd size and retention

Generally, we expect crowd size to lead to better performance when retention is low. Low retention should help alleviate many of the problems associated with increases in size. As we stated earlier, problems associated with increases in crowd size include coordination difficulties, conflict, and social loafing. Many of these problems in the organization literature have been associated with the addition of employees who are not a good fit with the organization (Kristof-Brown, Zimmerman, & Johnson, 2005). These employees can often be the source of conflict and unproductive interactions with others (Krackhardt & Porter, 1985). The removal of crowd members who are not a good fit should reduce some of the problems associated with increases in crowd size.

Low retention can represent a filtering process where members who are best able to contribute to the team and conform to the norms of the crowd stay while others leave. When new editors join crowds they often encounter a hostile environment and have to learn quickly whether and where they fit into the crowd (Halfaker et al., 2011; Halfaker et al., 2013). This environment could create a self-selection process, helping editors determine which crowd best fits their interests and skill set. Indirect evidence from the

organizational literature supports this assertion. McEvoy and Cascio (1987), in a meta-analysis of 24 studies, found that voluntary turnover was lower among higher-performing employees than lower-performing employees. In other words, high performers were more likely to remain with the company. This could reflect a self-selection system, where employees who fit in and feel successful remain and those who do not move on. This is supported by prior literature that found that low retention can be the result of unhappy employees leaving, which can create a much better work environment for the remaining employees (Krackhardt & Porter, 1985). For example, it is possible that some conflicts and edit wars are resolved or avoided when some editors simply leave the crowd.

Taken together, prior literature on retention in organizations suggests that the filtering process represented by low retention can reduce the process losses normally associated with increases in crowd size. On the other hand, as size increase in crowds with high retention, coordination problems, conflicts, and social loafing are likely to bog down crowds.

H5) Crowd size is positively related to crowd performance as crowd retention decreases.

Method

We collected data from articles on films from Wikipedia's WikiProject Film community. For each article, we obtained a complete list of edits, including time of the edit and username of the editor. Nearly 350,000 editors contributed at least one edit to an article in our dataset. We also obtained a list of all Wikipedia articles that each editor contributed to. We chose this particular community for two reasons. First, most of the articles in the community have been evaluated for quality — we provide more details of this evaluation below. Second, we focused our study on a homogeneous set of articles to minimize variation in other dimensions such as breaking news, unexpected events, and controversial topics. In this way, we prevented our results from being driven by these exogenous factors.

Article quality

To measure the quality of Wikipedia articles, we used each article's class as assessed by the project's own community through a peer review process.[1] This measure has been used as a proxy for quality in Wikipedia studies (Kittur, Chi, & Suh, 2008; Kittur & Kraut, 2008; Liu & Ram, 2011; Stvilia, Twidale, Smith, & Gasser, 2005; Wilkinson & Huberman, 2007), and there is evidence that the quality class is related to the quality assessments of outside reviewers (Kittur et al., 2008). There are six classes, or quality levels, to which articles can be assigned. In descending order of quality, the article classes are *FA, GA, B, C, Start*, and *Stub*. Articles in the classes *Start* and *Stub* contain very little information and have very few editors. Because we were interested in studying active and relatively large crowds, we dropped articles in Start and Stub from the analysis. After this filtering, our data contained 4,317 articles. We assigned each article a digit from 1 to 4 corresponding to the C, B, GA, and FA classes, respectively. There were 123 articles in FA, 629 in GA, 476 in B, and 3,089 in C.

Interest diversity

We measured the similarity in topical interests of two editors on Wikipedia by the similarity of the articles they edited across Wikipedia. Given editor u, we let A_u be the set of articles u has edited on Wikipedia. For every pair of editors (u_1, u_2), we measured their Jaccard similarity as $J_{u1,u2} = \dfrac{A_{u_1} \bigcap A_{u_2}}{A_{u_1} \bigcap A_{u_2}}$. This measure indicates the overlap among the articles edited by u_1 and u_2, while controlling for the total number of articles that the pair edited. For each article a, we let P_a be the set of all pairs of editors of article a. We defined the *interest diversity* of an article a, TD_a, as one minus the average Jaccard similarity of all pairs of editors. That is $TD_a = 1 - \dfrac{1}{|P_a|} \sum_{(u_1, u_2) \in P_a} J_{u_1, u_2}$. The average interest diversity was

0.994, with standard deviation of 0.009. When TD_a is high, the editors of a tend to have low overlap in the set of articles they edit, making them a more topically diverse crowd.

Local workload diversity (LWD)

Not all articles split the workload among editors in the same way. Some articles are characterized by having a small set of editors that contributes most of the edits, while most other editors contribute a very small number of edits. Other articles have editors who split the work more evenly. To measure the extent to which the editors of an article have a diverse or uniform workload, we used the *Gini coefficient* of their edits. The Gini coefficient is a measure of inequality originally used to measure dispersion in a country's income distribution. In this context, it measures dispersion in the distribution of edits among editors. We let E_a be the set of editors of article a, and we let $W_a(e)$ be the number of times editor e contributed to article a. We defined the *Local workload diversity (LWD)* of article a as the Gini coefficient of the set $\bigcup_{e \in E_a} \{W_a(e)\}$. An article where all editors contribute a similar number of edits has a low LWD, while an article where a few editors produce significantly more edits than the rest has a high LWD. The average LWD was 0.528, with standard deviation of 0.144.

This measure was proposed by Kittur and Kraut (2008) as a measure of implicit coordination. We argue that this measure also serves as a proxy for a type of diversity. This is because an article with high LWD has different types of editors, some who contribute very little and some who contribute a lot. On the other hand, an article with low LWD has only one type of editor because all editors produce roughly the same amount of work.

Experience

We measured the extent to which editors are engaged in editing all types of Wikipedia articles. For each article a, we defined the *experience* of its editors as the mean number of edits each editor

contributed to Wikipedia articles other than a. The average experience was 2,809 edits, with a standard deviation of 995.

Global workload diversity (GWD)

We measured the diversity of workload in Wikipedia as a whole in a similar way to how we measured LWD. This measures whether an article has a combination of editors who are heavy contributors to Wikipedia in general as well as editors who focus mostly on a single or very few articles, or whether most editors have about equal outer engagement. We defined the *global workload diversity (GWD)* of an article as the Gini coefficient of the number of contributions to other Wikipedia articles by each editor. The average GWD was 0.649, with a standard deviation of 0.084.

Retention

To measure retention, we calculated the probability that an editor who was active on a given month was also active in a future month. More precisely, for each article a and each month m in the articles' activity period, we let $A_{a,m}$ be the set of editors who edited article a during month m. We let $AF_{a,m}$ be the set of editors in $A_{a,m}$ who edited the article any time after month m. The retention rate of article a is

defined as $R_a = \dfrac{\sum_{m \in M} AF_{a,m}}{\sum_{m \in M} A_{a,m}}$. The average retention rate is 0.166 with a standard deviation of 0.076.

Hence, on average, 16.6% of editors who are active on a given month come back to edit the article.

Crowd size

We measured crowd size by the log of the number of editors for each article. The average log of number of editors was 4.66, with a standard deviation of 1.22. The actual average number of editors per article was 198, with as many as 2,618.
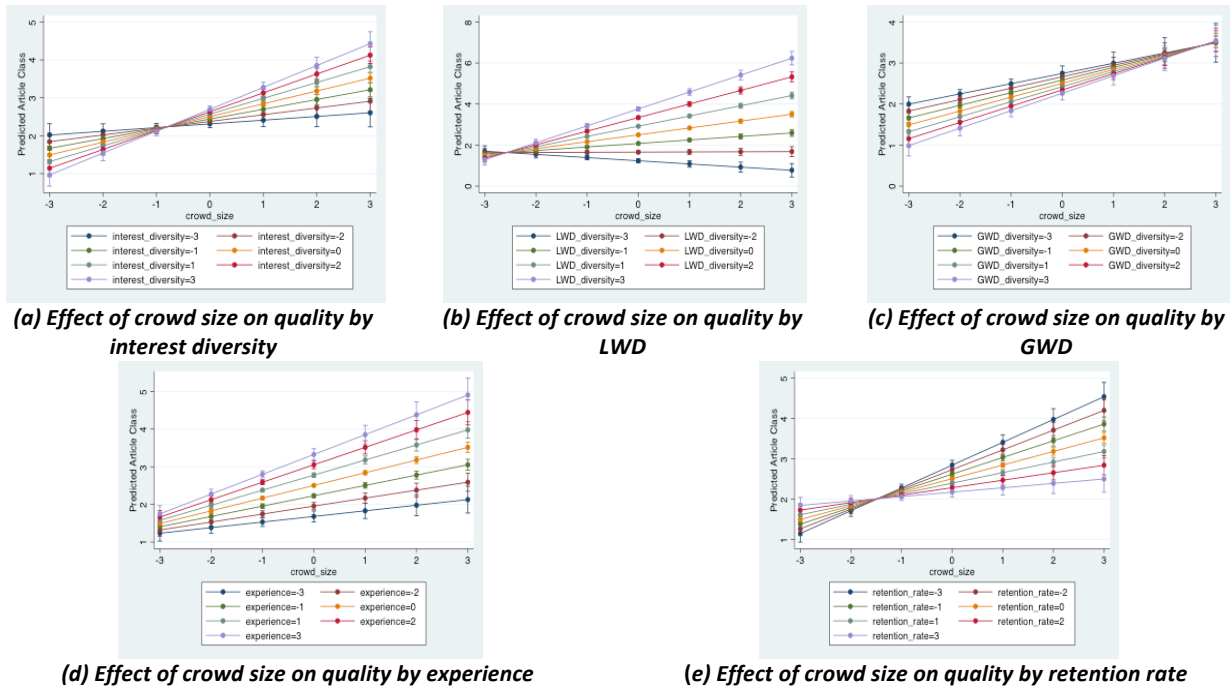
Results

We conducted a series of linear regressions to investigate the relationships among our variables. We used standardized independent variables in all regressions. First, we conducted linear regressions where article quality was the dependent variable. Table 1 and Figure 2 show the results from these regressions. In the main effects model, interest diversity, local workload diversity (LWD), crowd size, and experience were positively related to quality, and global workload diversity (GWD) and retention were negatively related to quality. The R^2 was 23%. We then included the moderation effects between all the independent variables and crowd size as recommended by Aiken and West (1991). All moderation effects were significant. The model which included the moderation effects has an R^2 of 26%. The increase in variance explained was significant at the 0.001 level.

Table 1. Results of regression analysis to predict quality from diversity measures, experience, retention, and crowd size. Model contains a total of 4,317 observations. Coeff = coefficients, SE = Standard Errors. Significance key: *p-value < 0.05, **p-value < 0.01, ***p-value < 0.001

Variable	Main Effects		Moderation Effects	
	Coeff.	SE	Coeff.	SE
Interest diversity	0.041**	0.013	0.064***	0.015
Experience	0.182***	0.016	0.274***	0.025
Global workload diversity (GWD)	-0.101***	0.020	-0.080**	0.029
Local workload diversity (LWD)	0.344***	0.014	0.421***	0.016
Retention rate	-0.034***	0.014	-0.112***	0.019
Crowd size	0.320***	0.019	0.335***	0.016
Interest diversity X Crowd size			0.080***	0.016
Experience X Crowd size			0.062***	0.014
GWD X Crowd size			0.029*	0.015
LWD X Crowd size			0.163***	0.014
Retention rate X Crowd size			-0.076***	0.013

Figure 2. The moderation effects of each measure and crowd size on quality. *GWD* global workload distribution, *LWD* local workload distribution

(a) Effect of crowd size on quality by interest diversity

(b) Effect of crowd size on quality by LWD

(c) Effect of crowd size on quality by GWD

(d) Effect of crowd size on quality by experience

(e) Effect of crowd size on quality by retention rate

These results support our hypotheses that crowd size is positively related to crowd performance in crowds with high diversity in LWD (H1a), GWD (H2a), and interests (H3a). This highlights that crowd size does not always have a positive effect on performance, as previously found in the literature (Chen et al.,

2010; Jehn, Northcraft, & Neale, 1999; Larson, Christensen, Abbott, & Franz, 1996; Page, 2007; Stasser, Taylor, & Hanna, 1994), and that diversity may be one of the factors that determine when crowd size is beneficial. The effect of experience is also greater for large crowds than for small crowds (H4a), which suggests that larger crowds benefit more from having editors who are overall more engaged in editing Wikipedia articles. Finally, the results also support our hypothesis that crowd size is positively related to crowd performance in crowds with low retention (H5), suggesting that as crowds become larger, the filtering out of workers may lead to increases in performance.

Next, we explored the impact of crowd diversity and size on retention. Table 2 and Figure 3 show the results of linear regression where retention is the dependent variable. The main effects model shows that interest diversity, LWD, and experience were positively related to retention, and GWD was negatively related to retention. This model has an R^2 of 35%. Crowd size was not significantly related to retention in the main effects model. However, when we added the moderation effects among all the independent variables and crowd size, we found that most of moderation effects were significant. Crowd size was positively related to retention in crowds high in GWD and interest diversity, supporting hypotheses H2b and H3b, respectively. However, we found no support for hypotheses H1b and H4b because crowd size was positively related to retention in crowds with low LWD, and the moderation effect between crowd size and experience was not significant.[2]

Table 2. Results of regression analysis to predict retention from diversity measures, experience, and crowd size. Model contains a total of 4,317 observations. Coeff = coefficients, SE = Standard Errors. Significance key: *p-value < 0.05, **p-value < 0.01, ***p-value < 0.001

| Variable	Main Effects		Moderation Effects	
	Coeff.	SE	Coeff.	SE
Interest diversity	0.003**	0.001	0.006***	0.001
Experience	0.005***	0.001	0.004*	0.002
Global workload diversity (GWD)	-0.015***	0.002	-0.007**	0.002
Local workload diversity (LWD)	0.041***	0.001	0.041***	0.001
Crowd size	0.002	0.002	0.000	0.002
Interest diversity X Crowd size			0.007***	0.001
Experience X Crowd size			-0.001	0.001
GWD X Crowd size			0.005***	0.001
LWD X Crowd size			-0.003**	0.001

Figure 3. The moderation effects of each measure and crowd size on retention.
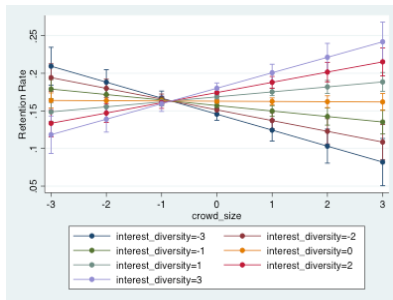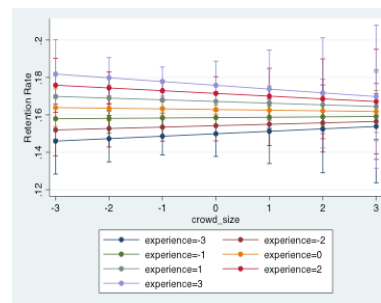

(a) Effect of crowd size on retention by interest diversity
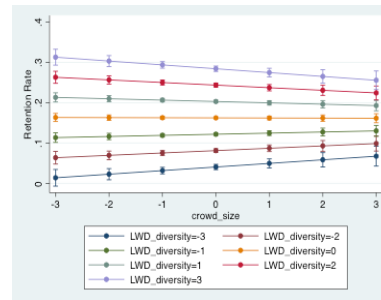
(b) Effect of crowd size on retention by LWD
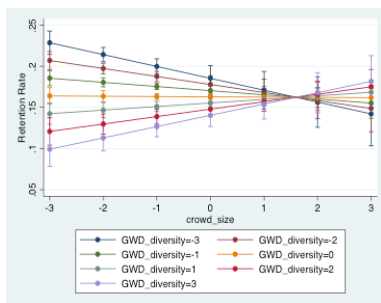
(c) Effect of crowd size on retention by GWD

(d) Effect of crowd size on retention by experience

Discussion

The goal of this paper was to examine the impact of crowd size on retention and performance in crowds. Specifically, we extended the research on crowds by demonstrating the role of crowd composition

21

on the relationship between crowd retention and performance. This research draws from the literature on crowds and on traditional organizational teams to better understand crowd retention and performance. Our results show that the influence of crowd size is contingent on other factors. Increases in size alone do not necessarily lead to better performance or retention. Thus, our results go beyond existing knowledge and contribute to the understanding of crowds. In the following sections, we discuss the contributions and implications of our research along with future research questions. Then we present the limitations associated with the study and our conclusions.

Contributions

Results of this study highlight that crowd composition is important to understanding when size leads to better performance. More experienced and diverse crowds can leverage increases in size to achieve better performance. The addition of individuals to crowds does not necessarily increase performance. Crowd size had little if any relationship with performance when local workload diversity (LWD), interest diversity, and experience were low. But, crowd size had a strong positive relationship with performance when these factors were high. The impact of global workload diversity (GWD) on the relationship between crowd size and performance was a little more nuanced and complex. Crowd size was positively related to crowd performance when GWD was low but the relationship became much stronger as GWD increased. This study confirms research that found a similar moderation effect of LWD on the relationship between crowd size and performance (Kittur et al., 2009; Kittur & Kraut, 2008) but also extends findings to include GWD, interest diversity, and experience.

Our findings also highlight the complex relationships among crowd size, composition, and retention. Crowd size can lead to more or less retention depending on the crowd's composition. Crowd size translates to better retention rates when crowds are high in both GWD and interest diversity. However, this was not the case for LWD. In fact, increases in size were negatively related to retention as LWD increased.

22

This was completely opposite of what we expected. Previous research on retention in Wikipedia crowds has used re-edits along with a lack of opportunities to participate in crowd work to explain low retention rates (Halfaker et al., 2011; Halfaker et al., 2013). LWD may embody both of these issues. Although LWD seems to be problematic for retention it is vital to helping crowds take advantage of their size to obtain better performance. If we assume that both better retention and better performance are our goals, it would seem that LWD presents a paradox. Crowd experience did not moderate the relationship between crowd size and retention. However, crowd experience was positively associated with crowd retention as a main effect. This indicates that experience is important to retention irrespective of the size of the crowd.

This study also contributes to our understanding of crowd retention and performance. Crowd size was positively associated with performance when retention was low but not when retention was high. The moderation effect coupled with the significant negative main effect of retention on performance seems to suggest that increases in retention do not benefit crowd performance. In fact, the opposite seems to be true. This provides evidence that crowds perform best when they are able to create an environment where members can find their place in the crowd or move on. Yet this is not without drawbacks. Crowds may be sacrificing long-term viability for short-term performance.

Finally, our results have important implications for collaborative systems that support crowd work. The use of crowds is becoming increasingly common across many domains, and this trend has inspired the design of recommender systems intended to maximize the benefits of large collaborations by suggesting new partnerships (Cosley, Frankowski, Terveen, & Riedl, 2007; McDonald, 2003). Currently, recommender systems use individual attributes like expertise and experience to make suggestions (Cosley et al., 2007). Our results suggest that recommender systems should also consider the number of editors and the diversity of the crowd along with how the new member would change both the size and diversity.

Limitations

This study has several limitations. First, the results of our study were found in one particular Wikipedia community. Future studies should be conducted to determine whether our findings hold true for other Wikipedia communities and for other communities outside of Wikipedia. However, our measures and the framework we utilized can be easily applied to crowds in other contexts. Therefore, it is feasible to validate our findings in other domains. Second, we examined three types of diversity. The literature on groups has identified many types of diversity and each could be important to understanding performance in crowds. Future research should be conducted to determine whether other types of diversity moderate the impacts of crowd size on retention and performance.

Third, although our results shed light on how the effects of crowd composition differ by size, we did not attempt to quantify the exact point at which crowd composition helps or hurts crowds relative to size. Based on our results this seems to depend on the type of composition attribute. For example, experiments could be conducted to determine exactly what amount of diversity, per type of diversity, translates to better or worse outcomes as crowd size increases. Further, the magnitude of the effects we found is not very large even when the effects are significant. This is expected given the relatively small sample size and because there are many other variables that affect performance and retention. We note that our aim is not to predict the performance and retention of a project with high accuracy, but to explore the relationships between the variables we considered. Finally, although we employed the literature on traditional organizational teams, there are significant differences. Traditional organizational teams tend to be much smaller than crowds and, unlike the crowds we studied, interact through face-to-face meetings. As such, it is unclear whether these results would hold true in traditional organizational teams.

Implications for research

This study has implications for research on crowd size and composition. Results from this study imply that the impacts of size are dependent on crowd composition and what makes larger crowds

successful is not the same as what makes smaller crowds successful. For example, we noticed a break-even point where smaller and less diverse crowds outperformed smaller and more diverse crowds. These smaller crowds performed better when they had much less crowd diversity and had higher retention. Larger crowds tended to perform better when diversity was high and retention was low. Yet, all crowds begin small and some grow while others die out. It is unclear how successful crowds make the transition from being small and homogeneous to becoming large and diverse. Future research should be conducted to determine which characteristics or behaviors enable or hinder crowds in making this transition. This research would be vital to understanding the lifecycle of crowds.

Going forward, scholars should begin to think more about the mechanisms by which the impacts of diversity relative to size influence retention and performance. Diversity is often referred to as a double-edge sword because it can have positive and negative effects on teamwork (O'Reilly, Caldwell, & Barnett, 1989). Diversity can be good for teamwork because it provides a source of unique ideas (Robert et al., 2008). However, it can be bad because diverse teams have greater difficulty working together effectively (Van Knippenberg & Schippers, 2007). Based on this premise, it appears that size is crucial to determining when diversity hinders the ability of group members to work well together. But more research is needed to fully examine why. In particular, future research on crowd diversity could explore mediators that could help explain why the moderation effect between size and diversity leads to better performance or more or less retention.

Although we caution against over-generalizing from one study, our findings call into question the value of retention relative to performance. The influence of retention on the relationship between crowd size and performance is both interesting and problematic. On one hand, retention seems to hurt the performance of crowds. On the other hand, the inability of one crowd to retain members may bleed over to other crowds. For example, if an editor quits one crowd how likely is that editor to remain in the Wiki community? Although the performance of a specific crowd may not be hurt in the short run, the lack of

available help for other less popular crowds could be hurt. Research should be conducted to determine whether newcomers simply move to other Wiki communities are leave altogether.

Conclusions

This study has examined the impacts of crowd composition on the relationship of crowd size, retention and performance. Based on our results, crowd composition is vital to determining when crowd size translates to better retention and performance. In addition, our results suggest that it is not always desirable to retain crowd members. Low retention rates were associated with better performance as crowd size increased. Overall, our results suggest that crowds benefit from increased size when they are diverse and experienced and have low retention rates.

Acknowledgments

References

Ahuja, M. K., Chudoba, K. M., Kacmar, C. J., McKnight, D. H., & George, J. F. (2007). IT road warriors: Balancing work-family conflict, job autonomy, and work overload to mitigate turnover intentions. *MIS Quarterly, 31(*1), 1–17.

Aiken, L. S., and West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions.* Newbury Park, CA: Sage Publications, Inc.

Alnuaimi, O. A., Robert, L., & Maruping, L. M. (2009). Social loafing in brainstorming CMC teams: The role of moral disengagement. *Proceedings of the 42nd Hawaii International Conference on System Sciences, Big Island, HI* (pp. 1–9). IEEE. doi:10.1109/HICSS.2009.396

Alnuaimi, O. A., Robert, L. P., & Maruping, L. M. (2010). Team size, dispersion, and social loafing in technology-supported teams: A perspective on the theory of moral disengagement. *Journal of Management Information Systems, 27*(1), 203–230.

Amason, A. C., Sapienza, H. J. (1997). The effects of top management team size and interaction norms on cognitive and affective conflict. *Journal of Management, 23*(4), 495–516.

Arazy, O. F. E. R., Morgan, W., & Patterson, R. (2006, Dec.). Wisdom of the crowds: Decentralized knowledge construction in Wikipedia. Paper presented at the 16th Annual Workshop on Information Technologies & Systems (WITS), Phoenix, AZ.

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* (pp. 233–236). New York, NY: ACM.

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*, 71–98.

Arazy, O., Yeo, L., Nov, O. (2013). Stay on the Wikipedia task: When task-related disagreements slip into

    personal and procedural conflicts. *Journal of the American Society for Information Science and*

    Technology (JASIST), *64*(8),1634–1648.

Armstrong-Stassen, M. (1994). Coping with transition: A study of layoff survivors. *Journal of Organizational*

    Behavior, 15(7), 597–621.

Bretz, R. D., & Judge, T. A. (1994). Person–organization fit and the theory of work adjustment: Implications

    for satisfaction, tenure, and career success. *Journal of Vocational Behavior, 44*, 32–54.

Butler, B. S. (2001). Membership size, communication activity, and sustainability: A resource-based model

    of online social structures. *Information Systems Research*, 12(4), 346–362.

Carillo, K. & Okoli, C. (2011). Generating quality open content: A functional group perspective based on the

    time, interaction, and performance theory. *Information & Management, 48*(6), 208–219.

Chen, J., Ren, Y., & Riedl, J. (2010). The effects of diversity on group productivity and member withdrawal

    in online volunteer groups. *Proceedings of CHI2010, Atlanta, GA.* New York, NY: ACM.

Choi, M. S., Woo, J. & Han, I. (2013). Understanding the group size for collaborative work in Wiki-based

    communities. *Proceedings of PACIS 2013.* Retrieved from http://aisel.aisnet.org/pacis2013/31

Cosley, D., Frankowski, D., Terveen, L., & Riedl, J. (2007). SuggestBot: Using intelligent task routing to

    help people find work in Wikipedia. *Proceedings of the 12th international conference on intelligent user*

    interfaces (pp. 32–41)*.* New York, NY: ACM.

Cox, T. H., & Blake, S. (1991). Managing cultural diversity: Implications for organizational competitiveness.

    The Academy of Management Executive, 5(3), 45–56*.*

Cummings, J. N., Kiesler, S., Zadeh, R. B., & Balakrishnan, A. D. (2013). Group heterogeneity increases

    the risks of large group size: A longitudinal study of productivity in research groups. *Psychological*

    Science, *24*(6), 880–890.

Daniel, S., Agarwal, R., & Stewart, K. J. (2013). The effects of diversity in global, distributed collectives: A study of open source project success. *Information Systems Research 24*(2), 312–333.

Dokko, G., Wilk, S. L., & Rothbard, N. P. (2009). Unpacking prior experience: How career history affects job performance. *Organization Science 20*(1), 51–68.

Gonzalez, J. A., & Denisi, A. S. (2009). Cross-level effects of demography and diversity climate on organizational attachment and firm effectiveness. *Journal of Organizational Behavior, 30*(1), 21–40.

Gotsis, G., & Kortezi, Z. (2015). Workplace diversity: A resource or a source of conflict? In *Critical studies in diversity management literature* (pp. 1–12). Dordrecht, The Netherlands: Springer.

Goyal, S., Maruping, L., & Robert, L. (2008). Diversity and conflict in teams: A faultline model perspective. In *Academy of Management Proceedings 2008*(1), 1–6.

Groysberg, B., Polzer, J., Elfenbein, H. (2011). Too many cooks spoil the broth: How high-status individuals decrease group effectiveness. *Organization Science 22*(3), 722–737.

Hale, D., Ployhart, R., & Shepherd, W. (2015). A two-phase longitudinal model of a turnover event: disruption, recovery rates, and moderators of collective performance. *Academy of Management Journal.* Advance online publication. doi:10.5465/amj.2013.0546

Halfaker, A., Geiger, R. S., Morgan, J., & Riedl, J. (2013). The rise and decline of an open collaboration community. How Wikipedia's reaction to sudden popularity is causing its decline. *American Behavioral Scientist.* 57(5), 664-688.

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 (WikiSym '11).* New York, NY: ACM.

Harrison, D. A., & Klein, K. J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review, 32*, 1199–1228.

Hausknecht, J. P., Trevor, C. O., & Howard, M. J. (2009). Unit-level voluntary turnover rates and customer

    service quality: Implications of group cohesiveness, newcomer concentration, and size. *Journal of*

    Applied Psychology, 94(4), 1068.

Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of

    diversity, conflict, and performance in workgroups. *Administrative Science Quarterly, 4*, 741.

Johnston, M. W., Parasuraman, A., & Futrell, C. M. (1989). Extending a model of salesperson role

    perceptions and work-related attitudes: Impact of job tenure. *Journal of Business Research, 18*(4),

    269–290.

Joseph, D., Kok-Yee, N., Koh, C., & Soon, A. (2007). Turnover of information technology professionals: A

    narrative review, meta-analytic structural equation modeling, and model development. *MIS Quarterly,*

    31(3), 547–577.

Kane, G. C. (2011). A multi-method study of information quality in Wiki collaboration. *ACM Transactions on*

    Management Information Systems, 2(1).

Kittur, A., Chi, E., & Suh, B. (2008). Crowdsourcing for usability: Using micro-task markets for rapid,

    remote, and low-cost user measurements. *Proceedings of CHI2008, Florence, Italy*. New York, NY:

    ACM.

Kittur, A., & Kraut, R. E. (2008). Harnessing the wisdom of crowds in Wikipedia: Quality through

    coordination. *Proceedings of the 2008 conference on computer supported cooperative work* (pp. 37–

    46). New York, NY: ACM.

Kittur, A., & Kraut, R. E. (2010). Beyond Wikipedia: Coordination and conflict in online production groups.

    Proceedings of CSCW '10 (pp. 215–224). New York, NY: ACM.

Kittur, A., Lee, B., & Kraut, R. E. (2009). Coordination in collective intelligence: The role of team structure

    and task interdependence. *Proceedings of the SIGCHI conference on human factors in computing*

    systems (pp. 1495–1504). New York, NY: ACM.

Krackhardt, D., & Porter, L. W. (1985). When friends leave: A structural analysis of the relationship between turnover and stayers' attitudes. *Administrative Science Quarterly, 30*(1985), 242–261.

Krieger, M., Stark, E. M., & Klemmer, S. R. (2009). Coordinating tasks on the commons: Designing for personal goals, expertise and serendipity. *Proceedings of CHI2009, Boston, MA* (pp. 1485–1494). New York, NY: ACM.

Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individuals' fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology, 58*(2), 281–342.

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

Larson, J. R., Christensen, C., Abbott, A. S., & Franz, T. M. (1996). Diagnosing groups: Charting the flow of information in medical decision-making teams. *Journal of Personality and Social Psychology, 2*, 315–330.

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), 1–23.

Long, B. C. (1993). Coping strategies of male managers: A prospective analysis of predictors of psychosomatic symptoms and job satisfaction. *Journal of Vocational Behavior, 42*(2), 184–199.

McDonald, D. W. (2003). Recommending collaboration with social networks: a comparative evaluation. *Proceedings of CHI2003, Ft. Lauderdale, FL.* New York, NY: ACM.

McEvoy, G. M., & Cascio, W. F. (1987). Do good or poor performers leave? A meta-analysis of the relationship between performance and turnover. *Academy of Management Journal*, *30*(4), 744–762.

Mesgari, M., Okoli, C., Mehdi, M., Nielsen, F. Å., & Lanamäki, A. (2015). "The sum of all human knowledge": A systematic review of scholarly research on the content of Wikipedia. Journal of the Association for Information Science and Technology, 66(2), 219-245.

Newell, J., Maruping, L., Riemenschneider, C., & Robert, L. (2008). Leveraging e-identities: The impact of perceived diversity on team social integration and performance. *Proceedings of ICIS 2008.* Retrieved from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1171&context=icis2008

Newell, J., Robert, L. P., Riemenschneider, C., & Maruping, L. M. (2009). Influencing individual perceptions of deep level diversity in virtual learning teams (VLT). *Proceedings of HICSS 2009.* IEEE Press. doi:10.1109/HICSS.2009.252

Ng, T. W. H., & Feldman, D. C. (2010). Organizational tenure and job performance. *Journal of Management, 36*(5), 1220–1250.

O'Reilly, C. A., Caldwell, D. F., & Barnett, W. P. (1989). Work group demography, social integration, and turnover. *Administrative Science Quarterly, 34*, 21–37.

Oliver, P. E., & Marwell, G. (1988). The paradox of group size in collective action: A theory of the critical mass. II. *American Sociological Review, 53(1),* 1–8.

Overbeck, J. R., Correll, J., & Park, B. (2005). Internal status sorting in groups: The problem of too many stars. In M. C. Thomas-Hunt (Ed.), *Research on managing groups and teams*: *Status and groups* (Vol. 7). Oxford, UK: Elsevier.

Page, S. E. (2007). *The difference: How the power of diversity creates better groups, firms, schools, and societies.* Princeton, NJ: Princeton University Press.

Qin, X., Salter-Townshend, M., & Cunningham, P. (2014). Exploring the relationship between membership turnover and productivity in online communities. *Proceedings of the eighth international AAAI conference on weblogs and social media, Ann Arbor, MI, USA, Association for the Advancement of Artificial Intelligence (AAAI),* (pp. 406–415).

Ransbotham, S., & Kane, G. C. (2011). Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in Wikipedia. *MIS Quarterly, 35*(3), 613.

Robert, L. P. (2013). A multi-level analysis of the impact of shared leadership in diverse virtual teams. *Proceedings of the CSCW* (pp. 363–374). New York, NY: ACM. doi:10.1145/2441776.2441818

Robert Jr, L. P., Dennis, A. R., & Ahuja, M. K. (2008). Social capital and knowledge integration in digitally enabled teams. *Information Systems Research*, 19(3), 314-334.

Robert, L., & Romero, D.M. (2015). Crowd Size, Diversity and Performance. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1379-1382). New York, NY: ACM.

Srinivasan, S. S., Maruping, L. M., & Robert, L. P. (2012). Idea generation in technology — supported teams: A multilevel motivational perspective. *Proceedings of the 45th Hawaii international conference on system science, Maui, HI* (pp. 247–256). IEEE. doi: 10.1109/HICSS.2012.308

Srinivasan, S.-S., Maruping, L. M., & Robert, L. P. (2010). Mechanisms underlying social loafing in technology teams: An empirical analysis. *Proceedings of ICIS 2010.* Retrieved from http://aisel.aisnet.org/icis2010_submissions/183

Stasser, G., Taylor, L. A., & Hanna, C. (1994). Information sampling in structured and unstructured discussions of three- and six-person groups. *Journal of Personality and Social Psychology, 57*(1), 67–78.

Stvilia, B., Twidale, M. B., Smith, L.C., & Gasser, L. (2005). Assessing information quality of a community-based encyclopedia. *Proceedings of ICIQ.* Retrieved from http://mitiq.mit.edu/ICIQ /Documents/IQ%20Conference%202005/Papers/AssessingIQofaCommunity-basedEncy.pdf

Suh, B., Convertino, G., Chi, E. H., & Pirolli, P. (2009). The singularity is not near: slowing growth of Wikipedia. *WikiSym*, *8*, 1–10.

Tesluk, P. E., & Jacobs, R. R. (1998). Toward an integrated model of work experience. *Personnel Psychology, 51*(2), 321–355.

Thomas, E. J., & Fink, C. F. (1963). Effects of group size. *Psychological Bulletin, 60*(4), 371.

Valacich, J. S., Wheeler, B. C., Mennecke, B. E., & Wachter, R. M. (1995). The effects of numerical and logical group size on computer mediated idea generation. *Organizational Behavior and Human Decision Processes, 62*(3), 318–329.

Van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annual Review of Psychology, 58,* 515–541.

Wagner, J. A., Ferris, G. R., Fandt, P. M., & Wayne, S. J. (1987). The organizational tenure–job involvement relationship: A job-career experience explanation. *Journal of Occupational Behavior, 8*, 63–70.

Weber, R. A., Camerer, C. F., Rottenstreich, Y., & Knez, M. (2001). The illusion of leadership: Misattribution of cause in coordination games. *Organization Science, 12*(5), 582–598.

Wilkinson, D. M., & Huberman, B. A. (2007). Assessing the value of cooperation in Wikipedia. *First Monday, 12*(4).

Windeler, J. B., Maruping, L. M., Robert, L. P., & Riemenschneider, C. K. (2015). E-profiles, Conflict, and Shared Understanding in Distributed Teams. *Journal of the Association for Information Systems*, 16(7), 608-645.

WMDW (2014). Charting diversity: Working together towards diversity in Wikipedia. Retrieved from https://upload.wikimedia.org/wikipedia/commons/5/57/Charting_Diversity.pdf

Wright, T. A., & Bonett, D. G. (1993). Role of employee coping and performance in voluntary employee withdrawal: A research refinement and elaboration. *Journal of Management, 19*(1), 147–161.

Wright, T. A., & Bonett, D. G. (2002). The moderating effects of employee tenure on the relation between organizational commitment and job performance: a meta-analysis. *Journal of Applied Psychology, 87*(6), 1183–1190.

Xiao, L. (2014). Effects of rationale awareness in online ideation crowdsourcing tasks. *Journal of the American Society for Information Science and Technology, 65*(8), 1707–1720.

Yu, L. L., & Nickerson, J. V. (2011). Generating creative ideas through crowds: An experimental study of

combination. *Proceedings of the thirty-second international conference on information systems.*

Abstract retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2039547

Notes

More details about the assessment process and quality classes can be found at
http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Film/Assessment
[2] Because more than 70% of the articles in our dataset were in quality class C, we repeated the regression analyses using only
quality classes B, GA, and FA. We observed that the trends were consistent in both sets.