Retrospective versus Prospective Evidence for Promotion: 
The Case of Wikipedia

Ben Collier
Moira Burke
Niki Kittur
Robert Kraut

Carnegie Mellon University

To be presented

Academy of Management Meeting, 2008

Abstract

Wikipedia is a large, online encyclopedia written and edited by volunteers. With a distributed work force of thousands of editors autonomously making changes, coordination in Wikipedia is a challenge. This paper examines how the process used to promote some editors to become administrators selects those who have the skills and temperament to act like managers in an organization without formal management roles. Studies of promotions suggest they have two complementary functions: (1) providing workers with incentives and rewards to deliver the effort and performance valued by the organization and (2) selecting workers for roles where they can contribute effectively to the organization. The incentive function leads organizations to promote based on retrospective indicators of past achievement, while the selection function leads them to promote based on prospective predictors of future job performance. We examined 1502 promotion decisions in Wikipedia to determine the influence of retrospective and prospective evidence for promotion. Two studies – a qualitative analysis of reasons cited in promotion decisions and a quantitative, policy-capture analysis relating users’ historical editing patterns to promotion success – both show that Wikipedia bases its promotion decisions on prospective evidence. These findings are consistent with the need in Wikipedia to recruit and select individuals without formal authority or explicit hierarchy to do the difficult job of coordinating the work of others. A comparison of the two studies, however, suggests that while Wikipedians talk about promoting editors based on their record of editing articles, data from the policy capture study indicates they do not, in fact, use these criteria. We discuss the implications of these findings for theories of promotion, policy capture, and online groups.
Introduction

Wikipedia is a large, open source, online encyclopedia written and edited by volunteers. As of December 2007 over 75,000 editors had contributed over 2.1 million articles to the English version alone. Smaller but active Wikipedia versions exist in German, French, Japanese, Polish and other languages. Although Wikipedia is written and edited by volunteers and is not supervised by a professional staff, evidence suggests that the quality in Wikipedia is comparable to that of the Encyclopedia Britannica (Giles, 2005). Although organizational scholars have asked why volunteers contribute to Wikipedia and similar environments where unpaid volunteers contribute content, they have written less about how these highly decentralized environments coordinate work to produce high quality products (Bryant, Forte, & Bruckman, 2005; Roberts, Hann & Slaughter, 2006). Wikipedia as an organization is substantially less hierarchical than a conventional production company. By contrast, Encyclopedia Britannica has a CEO and other corporate officers, an editor-in-chief, a scholarly advisory board, and a staff of 19 full-time editors who vet and supervise the work of over 4000 contributors.

To help coordinate work in Wikipedia in the face of a large and largely unsupervised workforce, Wikipedia has developed an elaborate system of policies that lay out Wikipedian values such as civility and neutral point of view and work procedures such as those for deleting pages, resolving disputes and electing administrators (Butler, Joyce, & Pike, 2008). It also empowered a class of administrators with special tools to enforce some of the policies. For example, these administrators can ban other editors from working in the encyclopedia, protect pages from changes and often serve as formal and informal mediators in resolving disputes. Administrators in Wikipedia differ from regular editors in two main ways. First, they have additional tools for page deletion and user maintenance, and regular editors request assistance from administrators at
a public notice board for such issues as vandalism, page protection, inappropriate usernames, conflict between users, and page deletion. Second, administrators have elevated social status, as trusted editors who understand policy and its application. Though their position is ostensibly a lowly, janitorial one (as symbolized by their logo of a bucket and a mop), administrators in practice are well regarded as acting in the community’s best interest and having enough experience to use their tools wisely. Approximately 2700 editors have been nominated for administrator status since 2001 with an overall promotion success rate of 53%. However, the process has gradually grown more rigorous, dropping from a 75.5% success rate through 2005 to 42% in 2006 and 2007, and some early administrators have expressed doubt that they would pass muster if their promotion debate were held today (Forte & Bruckman, 2008).

According to Milgrom and Roberts (1992), promotions serve two roles in an organization. First, they are part of a reward and incentive system to induce employees at one level of an organization to work hard in the organization's interest so that they can be promoted to positions of more power and responsibility. Second, they are part of a selection process in which people are vetted and assigned to the roles where they can best contribute to the organization's performance. This paper examines the criteria that Wikipedia uses to promote regular editors to administrator status and the way in which the relatively flat organizational structure and large, decentralized workforce in Wikipedia influence these criteria.

Practitioners, organizational scholars and economists alike agree that rational contest models of promotion, which focus on the quality of candidates’ prior job performance, are a partial explanation for who will be promoted (Turner, 1960). In a contest model, candidates at one level of a corporate hierarchy compete against each other for promotion to the next level based on
their job performance in the lower position. "What makes the greatest difference in getting ahead . . . is performance on the job and adding value . . . People compete with each other in an open and fair contest for advancement, and victory comes to those who demonstrate the greatest accomplishments" (Ng, Eby, Sorenson, & Feldman, p. 369). Contest models are consistent with common organizational practices of counting sales commissions, journal publications or 3-point shots as criteria for promotion among sales organizations, professors and National Basketball Association players, respectively. Scholars often use these rational, contest models as the foil when discussing models of promotion that appear less rational and that emphasize factors less relevant to job performance, including height (Melamed & Bozionelos, 1992), good looks (Hamermesh & Biddle, 1994), gender (Maume, 1999) or social ties (Burt 2001) or other types of sponsorship (e.g., Hargen & Hagstrom, 1967).

Whether promotion functions as an incentive or a selection mechanism is likely to determine the types of evaluation criteria managers use in making promotion decisions. To the extent that promotions serve as incentives, managers should look for evidence that candidates for promotion have performed well in the past, independent of the degree to which prior job performance predicts performance in a subsequent one. Because the promotions are being used to reward past behavior and encourage effort and work quality that the organization judges as valuable, managers should base promotions on the type of behavior the organization wants to encourage, whether or not this behavior would serve the organization or candidate when the candidate ascends to a new position. For example, in a production environment like a factory, managers should promote the most productive workers to supervisor positions. In this sense, the criteria that managers use to make promotion decisions should be retrospective or backward-looking; they should be looking for evidence of good performance in the prior job.
This retrospective use of evidence is a partial explanation for the "Peter Principle," in which employees rise to their level of incompetence (Peter & Hull, 1969). People who are good at one job may not necessarily be good in the job into which they are promoted. If promotion candidates are being judged on a set of behavior, such as their individual productivity, for a position that requires a different set of knowledge, abilities, and skills then they will not necessarily be well qualified for the new position. This might be the case, for example, when professors are promoted to be department heads based on the quality of their research or basketball players promoted to be coaches based on their athletic talents; the skills needed for the more managerial positions of department head or coach may have had little to do with the candidate’s successes as academic or athletic superstars.

In contrast, to the extent that promotions serve as a selection device to identify appropriate candidates for higher-level, more difficult or more valuable organizational roles, then managers should use prior job performance primarily as evidence of future performance in these new roles. In this view, the promotion view is a prediction task, and the promotion criteria that managers use should be prospective or forward-looking, rather than retrospective, as they try to use prior job performance to predict performance in the new job.

Retrospective and prospective criteria for promotion are not necessarily mutually exclusive; in many cases some behavioral criterion for promotion may serve both as an incentive to encourage incumbents in the current job and to predict performance in a future one. Thus behaviors that indicate conscientiousness or willingness to work hard may generalize well, and using them as promotion criteria is valuable both as an incentive and as prediction. Many human-capital factors commonly used as criteria for promotion—such as work ethic or educational level (e.g., Ng et al,
have both functions. However, other potential promotion criteria, such as those indicating individual productivity in a production task (e.g., papers published in academia or points scored in basketball) may not be relevant to promotion to a more managerial position.

In this paper we use data from over 1500 promotion decisions elevating regular editors in the online encyclopedia Wikipedia to administrator status. Wikipedia is an interesting and important organizational phenomenon in its own right. Moreover, because all the work done on the encyclopedia and all the communication among participants is archived, Wikipedia makes visible organizational processes that are difficult to observe in conventional work organizations. Here we examine the process in Wikipedia in which ordinary editors become administrators in order to distinguish the use of retrospective and prospective evidence in promotion.

Our primary research goal is to understand the type of evidence judges use to promote editors to administrator status. The workforce in Wikipedia is largely autonomous and distributed. Individual editors decide which articles to work on and how much to work. There is no employment contract binding the volunteers to the organization and no supervisors assigning them tasks. Under these conditions one might expect that Wikipedia needs to use promotion as part of a reward and incentive system to motivate editors to do the needed production work. If this is the case, then Wikipedia should promote people who are highly productive editors -- contributing to many articles and projects and having discussions with other Wikipedians about these editing tasks. That is, if Wikipedia is using promotions primarily for purposes of rewarding and encouraging desired behavior on the part of editors, then we'd expect that the criteria for promotion in Wikipedia should primarily be retrospective and production oriented. However, given that Wikipedia has a very large population of volunteers who already do
production work without concrete rewards, promotion to reward production work might not be especially valuable to the organization.

On the other hand, there is much managerial work such as coordination and conflict resolution that needs to be accomplished in Wikipedia, with relatively few formal mechanisms to do so. If Wikipedia is using the promotion process to recruit and select individuals to do the difficult job of coordinating the work of others without formal authority or explicit hierarchy, then it should use more managerial-like criteria for evaluating candidates for promotion. The goals should be to promote people whose behavior as editors signals that they can become good administrators. Thus, the decision-makers should be looking for evidence that the candidate has the breadth of experience to handle the unanticipated demands made of administrators, the interpersonal skills needed to be civil while still handling the conflict frequently encountered in Wikipedia, and the temperament and interests to deal with Wikipedia policy.

We base our observations on two types of data. First, we examine formal Wikipedia policy and the standards members of the community explicitly invoke in their debates with each other about whom to promote to administrator status. These linguistic data provide evidence about the criteria the community claims to use. Second, because people often possess poor insight into their own evaluation policies, we use a policy-capture technique (Hobson & Gibson, 1983) to examine the details of Wikipedians’ behavior that predicts whether they will be selected to be administrators or not.

**Promotion: The Request for Adminship Process**

To become an administrator, an editor must undergo a week of scrutiny known as the Request
for Adminship (RfA), during which the community builds consensus about the candidate’s experience and trustworthiness. Administrator tools are not granted lightly; an inexperienced, biased, or ill-intentioned editor could cause significant damage, reducing the encyclopedia’s credibility or demotivating other editors. Thus, RfA contributors—those who voice support or opposition to the candidate’s adminship—examine a broad range of evidence including a strong edit history, diverse experience, polite interaction with other users, policy understanding, conflict resolution, and high quality work.

The process consists of three parts: an introductory nomination statement, the nominee’s answers to questions about past and future behavior, and statements of support, opposition, or neutrality by community members. Any registered Wikipedia member can voice an opinion, but the RfA is not a strict vote: at the end of a week, a bureaucrat—one of approximately 25 editors with privileges greater than administrators—reads through the opinions and decides whether consensus was reached. Candidates with more than 75% support are generally successful, though bureaucrats weigh voters’ reasons, not just their votes. Votes by suspected sockpuppets (multiple identities held by the same person) or meatpuppets (new users recruited by a voter to back up the voter’s opinion) are discounted. Nominees may withdraw at any time, and the “snowball clause”—a derivative of Wikipedia’s famous “ignore all rules” rule—allows any editor to close a nomination early to avoid wasting the community’s time if the nominee is so inexperienced as to not stand a “snowball’s chance in hell” of passing.

RfA candidates answer three standard questions as part of their nomination: (1) What chores do you intend to help with? (2) What are your best contributions, and why? and (3) Have you been involved in conflicts over editing or have other users caused you stress? Nominees answer these
and any additional questions posed by other community members, citing records of past events to demonstrate their competence and handling of controversy.

While RfA contributors have differing standards, many look for an answer to the first question that demonstrates the candidate is already behaving like an administrator, helping with chores such as monitoring recent changes, welcoming newcomers, or participating in debates at Requests for Comment (RfC) or Articles for Deletion (AfD). Editors are encouraged to first help the wiki using their existing privileges, such as reverting vandalism, organizing collections of articles in a subject area (WikiProjects), and reducing the non-administrator backlog, before considering promotion to administrator status. Only when new privileged abilities are needed are users encouraged to seek adminship, such as if a diligent vandal fighter is hampered in fighting vandalism by frequently waiting for administrators to block users she identifies.

The second RfA question demonstrates the significance of a candidate’s contribution, and many successful candidates discuss their work on articles that reached high levels of quality, such as Featured Article status. Answers to the third question highlight incidents in which the nominee has dealt with interpersonal conflict, a common occurrence in an encyclopedia edited by thousands of people with differing viewpoints yet striving for neutrality. Previous work has shown conflict on even seemingly neutral topics like chocolate, and that administrators often serve as unofficial mediators in edit wars on more controversial topics, such as euthanasia and evolution. In answering the third question, the nominee demonstrates how he has dealt with controversy or uncivil comments from other editors, linking to evidence on talk pages. Candidates involved in heated “edit wars” are unlikely to be well received by the community, though candidates who make full disclosures of previous mistakes and have recent histories of
good behavior are more likely to succeed.

**Study 1: Explicit Mentions of Criteria for Promotion**

**Rationale**

In this study we used Grounded Theory methods to examine the criteria cited by voters in promotion decisions (Glaser and Strauss, 1967). During the RfA process voters are strongly encouraged to leave a comment as to why they voted the way they did. By characterizing these comments we aim to understand whether administrator promotion is prospective or retrospective in nature. A major advantage of examining this question in the context of Wikipedia is that it provides us with complete records of an actual promotion decision context, rather than having to rely on laboratory or other artificial settings which may bias results (Hobson & Gibson, 1983).

**Method**

We randomly selected 25 successful and 25 unsuccessful administrator candidates from the period of November 2005 to May 2007. Evaluators were blind to the criteria in the guide to the RfA and those outlined by current administrators. A preliminary analysis of each candidate revealed several salient criteria mentioned by voters for both successful and unsuccessful candidates. In some cases, the criteria was mentioned as a reason the candidate should not be promoted (Negative Comment), in others it was mentioned as a reason the candidate should receive the promotion (Positive Comment). These criteria were categorized into either Retrospective criteria that demonstrated the candidate’s abilities and experience as an editor or Prospective criteria more suggestive of skills and personal characteristics required of an Administrator. Within the Prospective category, the data was further divided into aspects that
relate the candidates (1) Personal characteristics such as their knowledge, personal interaction, and trustworthiness and (2) Task characteristics which relate to work the candidate has done that strongly relates to work an administrator would perform.

**Retrospective Criteria**

*Number of Edits* is perhaps the simplest and most salient characteristic voters have available to them when evaluating an RfA. The community often discusses edits in terms of a necessary, but not sufficient condition to becoming an administrator. Voters suggest that edits should be "sufficient" (P37, 1) or "above the threshold" (P6, 52) before considering other factors.

*Time in Wikipedia* relates to how long the candidate has been a registered editor of Wikipedia. Candidates who have been a part of Wikipedia for an extended period of time are often noteworthy as it shows both a dedication to continue working and a sense that this candidate is ready for promotion. One voter comments, "[the candidate has] served his time, has stayed around to help the project, and does good work" (P26, 82).

*Article Quality* describes mentions of featured articles the candidate may have worked on, or other mentions of producing high quality article work.

**Prospective Personal Criteria**

*Knowledge of policy* describes whether the voter perceives the candidate understands the policies, guidelines, and processes that govern the Wikipedia community. While Wikipedia may on the surface appear to be ad hoc, thousands of policies are in place to govern the production of articles and the community (Butler, et al., 2008). Administrators are one of the key enforcers of
policy and must know how to intervene in disputes; in addition, administrators play a key role in
discussing and creating new policies to confront new challenges. A voter who provides a positive comment regarding a candidate’s knowledge of policy states, “In my experience with [candidate], I found him to be reasonable in confronting POV-pushing editors. I believe [candidate] has demonstrated sufficient knowledge of Wikipedia policy to wield the tools” (P25:81). Conversely, those who are clearly not up to snuff on Wikipedia policy may receive an oppose vote that comments “[candidate] is an excellent editor but I don’t think you are hot enough on policy yet” (P10, 38).

*Interaction with others* encompasses how voters comment on their communications with the candidate in Wikipedia. Whether they were working on an article together, discussing a project, or giving pointers on how to improve their style users often comment on whether the person was polite, civil, friendly, or helpful. Examples of a positive comment on a candidate's interpersonal skills include, "I have had frequent interaction with [candidate], and he's extremely polite and civil" (P9, 126). If a candidate has a history of uncivil interactions with users since Wikipedia is archival in nature it is available for all to see and will often not go unnoticed. "I get a bad feeling about this candidate's temperament and skills at dealing with other editors. Seems a little to prone to snap at others and be a little blunt" (P20, 086).

*Trustworthy* candidates have somehow demonstrated a history in Wikipedia that voters deem that unlikely to abuse the administrative tools or status. After an examination of possible faults in the candidate such as uncivil behavior, edit waring, or vandalisms, voters often comment they have found no cause for concern in this candidate. One voter's only comment on a supportive vote states: "Everything checks out, trustworthy" (P14, 62).
Neutral point of view is one of the key policies in Wikipedia and a necessary quality for anyone mediating conflict. Voters to describe someone as "fair" (P4, 63), "level-headed" (P9, 12), and "calm and neutral" (P16, 121) when a candidate has demonstrated they do not take sides on a topic but rather works to build consensus.

Conflict management relates strongly to a candidates demeanor and interpersonal skills in negotiating through difficulty. Often two parties have differing opinions on what an article should say or how something should be done and administrators need to have a calm, reasoned, and friendly mediation to resolve conflict. Often the voters who comment on a candidates behavior have personally interacted with the candidate and had negative or positive experiences with the candidate’s conflict management skills. Comments such as “helpful in our conflict” (P4, 25) and “very good mediator” (11, 152) are interpreted as positive comments, while the comments such as “I am concerned about [the candidate] being able to keep a cool head in confrontational situations” (p24, 221) suggest the candidate has poor conflict management skills.

Community involvement criteria that voters consider when determining if the candidate is not only editing articles but becoming involved in the Wikipedia community at large. Often voters will acknowledge the candidate has sufficient editing history, but note that "admins ought to be active within the community" (P7, 150). An ideal candidate for Adminship according to some voters should be "a well rounded user with heavy community involvement" (P53, 16).

Commitment to Wikipedia is a rather abstract criterion that voters perceive based on a number of candidate features such as tenure in Wikipedia, steady involvement, and consistency in their interactions. Voters are looking for candidates who are "clearly committed to Wikipedia in a
long-term way" (P13, 21).

Prospective Task Criteria

*Need for admin tools* describes criteria relating to whether the user has demonstrated they have previously worked on tasks that would benefit from having the Administrator tools. If a user has submitted requests for page moves or has demonstrated work in fighting vandalism, voters comment that the candidate "could do even better things to help with admin tools" (P37, 53).

*Vandalism* to Wikipedia pages includes deletion of pages by users and other defamation of Wikipedia content. In addition to editing, some users have taken it upon themselves to watch a set of pages to ensure users with ill intent do not destroy valuable content, and thus become "experienced vandal fighters" (P11, 80).

*Work in deletion voting* shows that candidates are working in a forum for unneeded articles, categories, and templates that are voted on to be deleted. Work in this area is characteristic of work an administrator will do once they have the position, but voting and discussion surrounding content to be deleted reflects positively as it demonstrates familiarity with administrative tasks.

*Edit diversity* reflects the extent to which a candidate is working on various sections of Wikipedia rather than only contributing to articles. Voters prefer candidates who have experience working on projects, discussing policy, and interacting with other users. Candidates who have "made excellent contributions across several namespaces" (P47, 38) are likely to have a variety of experience to draw from when they become administrators.
Results

To empirically summarize the results this study used the number of unique counts of a comment category per candidate, for example, for a given RfA if a handful of voters all mention that this candidate has a strong history of fighting vandalism the results show only one unique comment on positive mention of vandalism work. If in the same RfA several voters mention that someone had positive (friendly or civil) interaction with the candidate, and another set of voters comment that they had a negative (harsh or uncivil) interaction with a candidate, the results will show that the candidate had one unique positive interaction, and one unique negative interaction. By counting unique comments per candidate we eliminate a skew in the data that results from having one highly qualified or highly unqualified candidate have dozens of comments for a given criteria such as when a user clearly has a low number of edits and dozens of voters point out that they are opposing due to a lack of edits. Figure 1 shows the aggregate counts of how many RfA's cite each criteria.

---Insert Figure 1 about here---

Prospective task criteria. Candidates who had not sufficiently shown a need for the Administrative tools were heavily noted by voters and punished for their lack of participation tasks characteristic of administrators. This suggests evaluators have a forward-looking view of candidates, and want to know candidates for promotion are already beginning to work on tasks that they will undergo once promoted. Candidates with a strong history of fighting vandalism in Wikipedia received notable attention for their efforts, but those who did not were rarely punished. The diversity of candidates work and their participation in deletion discussion were mentioned in roughly a third of the candidates primarily in a negative light. Those who lacked
diversity or participation were noticed and called out for their shallow experience.

Prospective personal criteria. Candidates' knowledge of policy was the most widely discussed feature during their evaluation, with both high discussion of positive and negative aspects of their knowledge. A majority of the negative comments came from evaluation of candidates were not promoted to be an Administrator. This suggests that regardless of other criteria, if the candidate does not have the cognitive knowledge and has not demonstrated their familiarity with policy they do not have the personal qualities necessary to be promoted. Throughout a candidate’s history voters evaluate the candidate’s interpersonal skills and trustworthiness as next two most cited criteria. Voters appear to be keenly aware that upon promotion editors will be in the spotlight to other users and have more frequent interactions with other administrators; promoting someone who is a strong worker but difficult to get along with or untrustworthy does not benefit the community. As expected, a candidates apparent neutral point of view, conflict management abilities, and commitment to Wikipedia are all widely cited as criteria for promotion. Additionally it is interesting to note that when considering policy knowledge, conflict management, and community involvement voters cite a fairly even split between positive and negative comments perhaps due to the subjectivity of the evaluation process and variation in the personal experiences with that particular candidate.

Retrospective criteria. Voter frequently discuss the raw number of edits a user has, perhaps because that information is so readily available and is thought to be a general proxy for experience and worth ethic. Tenure within Wikipedia and article quality are of relatively little consideration when compared to other criteria.

The results summarized in Figure 1 show that a vast majority of voting behavior (83% of unique
category votes) reflects a Prospective look at Administrator promotion. Within the Prospective
category, a large portion of the votes relate to personal characteristics of the candidate that make
them fit for the new Administration position. Having a strong history of editorial behavior is still
highly considered, but these results would suggest that they are a necessary but not sufficient
condition for promotion. The four most frequently cited criteria are all within the prospective
view of promotion. The Wikipedia voting community appears to regard the promotion to
Administrator as a matching of finding the most qualified person for the position in terms of
personal factors and skills rather than as a reward for hard work or longevity within the
community.

**Study 2: Modeling Successful RfA Candidates**

*Rationale*

The analysis of reasons cited by voters for promotion decisions in Study 1 suggests a focus on
prospective rather than retrospective factors. However, previous research on policy capture has
found a disconnect between the factors people cite for making decisions and the factors actually
promotion in Wikipedia has a number of characteristics ideal for policy capture research: the role
of the Wikipedia administrator is well-defined; there are many judges involved in the promotion
decision; and there is high transparency in the records of past action since all edits are recorded.
In Study 2 we examine the link between quantitative prospective and retrospective factors of
candidate performance and promotion success. We also compare the factors cited in Study 1 as
important to promotion success with the factors actually used in promotion decisions.
Methods

The data include all 1551 Requests for Adminship from January 2006 through October 2007, with 49 RfAs removed for being multiple attempts by the same candidate in one month (all of which failed), bots, sockpuppets (multiple identities held by the same person), or because the nominee’s edit history prior to the RfA was not available. For each RfA, data from the user’s contribution history page up to the month before the RfA was parsed and counted, and grouped into the categories introduced in Study 1. Three categories from Study 1—Trustworthiness, Commitment to Wikipedia, and Article Quality—were not included in this analysis because they could not be captured with simple edit counts. These categories are discussed further in the discussion section. Table 1 provides summary statistics for the features.

Retrospective criteria

Number of edits includes the number of edits to articles and to WikiProjects.

Time in Wikipedia is the number of months between a user's first edit and his or her RfA.

Prospective task criteria

Edit diversity includes a count of the number of different areas in which the user has participated, from the set \{article, article talk, Wikipedia, Wikipedia talk, user, user talk, articles/categories/templates for deletion (XfD), (un)deletion review, other RfAs, village pump, administrator intervention against vandalism (AIV), requests for protection (RfP), administrator noticeboard, arbitration committee, mediation committee, and wikiprojects\}. So, a user who has edited articles, her own user page, and posted once at the Village Pump would have a diversity score of 3. Actual number of edits in each of these sections is accounted for in the following
categories to determine their relative importance.

*Need for admin tools* includes instances in which the editor requested help from administrators to perform a chore not available to normal editors. It is a count of posts to the administrators' noticeboard (e.g. for reporting inappropriate usernames or edit wars).

*Vandalism* includes reversion of vandalism (noted by “revert” or “rv” in the comment text accompanying an edit), requesting administrator intervention for specific vandals (AIV), and requesting protection for particular pages from malicious editors.

*Work in deletion voting* includes participating in deletion discussions, including articles/categories/templates for deletion (XfD), (un)deletion reviews, and votes for deletion.

**Prospective personal criteria**

*Knowledge of policy* includes edits to the Wikipedia namespace, a section of the encyclopedia devoted to policy and guideline discussions. Edits to WikiProjects, which are also part of the Wikipedia namespace, have been separated and counted below.

*Interaction with others* includes edits to "talk" pages. Every piece of content on Wikipedia (including articles, user profiles, and policies) has an associated "talk" page, on which changes to the content are discussed. Simple politeness markers, including any variant of “please” (including “pls” and “plz”) or “thanks” (including “thx”) in comment text are also counted.

*Neutral point of view* includes the number of edits in which neutrality fixes were made, noted by “(n)pov” for “neutral point of view” in the edit summary text.
Conflict management includes the number of posts to the arbitration or mediation committee pages, or "wikiquette" alerts (an early stage in dispute resolution).

Community involvement includes participation in other editors' Requests for Adminship, posts to the community discussion board known as the Village Pump, and welcoming posts to newcomers' talk pages.

Results

To examine the impact of these behavioral factors on the likelihood of a candidate's promotion, we performed a probit regression on the binary dependent variable, RfA success. All variables were standardized, so the coefficients in Table 1 represent the change in probability of success when a continuous variable is increased by one standard deviation. All variance inflation factors (VIFs) were well below 10, indicating low multicollinearity between factors. Multiple attempts by the same candidate in a single month were excluded, leaving only one attempt per month, and the candidate’s number of previous RfA attempts (in other months) is included as a control variable; each subsequent attempt has a 14.8% lower chance of success than the previous one. The model is 74.8% accurate in classifying RfA attempts as successful or not.

--- Insert Table 1 about here ---

The results strongly support the notion that mere retrospective edit counts ("editcountitis") are poor predictors of promotion in Wikipedia. Instead, prospective criteria, particularly interpersonal ones, are more influential in Request for Adminship decisions.

Prospective task criteria. Overall, the editor's diversity of experience in Wikipedia had a strong
impact on the likelihood of promotion. Branching out to additional regions demonstrates an understanding of the encyclopedia and its community as a whole, and editors with experience in additional sections were more likely to become administrators. Editing approximately 4 new regions was associated with an 8.6% increased likelihood of promotion. Reverting vandalism tended to be associated with an increased likelihood of promotion, as well. However, demonstrating a need for the tools by elevating issues to administrators' attention tended to be associated with a decreased likelihood of promotion: posts to the administrator noticeboard modestly decreased chances, and requesting administrator intervention against vandalism or for page protection were non-significant but in the negative direction. Finally, participation in page deletion discussions had no impact on promotion.

**Prospective personal criteria.** Participation in policymaking discussions is a strong predictor of RfA success. As administrators are expected to be well versed in policy and its application, this is consistent with a prospective decision making process. Instances of user interaction--"talk" pages--tell a particularly interesting story. The norm in Wikipedia is to hold discussions about article content on article talk pages, but these edits do not predict promotion, and neither do edits to Wikipedia namespace talk pages. However, when disagreements over content covers multiple articles or a user's overall behavior, the discussion often moves to user talk pages. Thus, user talk edits are likely to be mixed, and may have higher interpersonal conflict. As a result, user talk edits strongly decrease the likelihood of promotion. This is supported by the finding that edits to Arbitration or Mediation Committee pages, or to Wikiquette notices, all of which are venues for dispute resolution, decreases the likelihood of success, as well. Thus, successfully promoted Wikipedia administrators appear to avoid conflict-related interpersonal behavior.
Politeness helps modestly; though it was rare, every 40 instances of saying “thanks” in comments increased the likelihood of success by 7.6%. Saying “please” had no effect. Upholding Wikipedia's much valued neutrality by fixing point-of-view issues was also associated with a greater likelihood of promotion. Somewhat surprisingly, involvement in the community through welcoming newcomers, participating in other RfAs, or chatting at the Village Pump also had no effect.

*Retrospective criteria.* Merely editing large numbers of articles was insufficient for promotion in Wikipedia. Length of experience helps slightly; every 7 additional months between a user's first edit and her RfA increased the likelihood of success by 3.4%.

**Discussion**

We performed two studies investigating the use of prospective and retrospective factors on administrator promotion decisions in Wikipedia. The first study characterized the reasons cited by voters in their selection decisions, finding a strong influence of prospective factors including policy knowledge, interaction with others, trustworthiness, and a neutral point of view. Fewer retrospective factors were cited, although total past edits was frequently mentioned.

The second study involved a quantitative analysis relating candidates’ actual prior experience to promotion success. We found a number of prospective metrics significantly correlated with promotion success, including experience with policies and interaction with others. Fewer retrospective factors were relevant to success, although tenure in Wikipedia was found to be significant.
The prospective metrics cited in the first study were largely consistent with those found to be actually predictive in the second study. However, there were also differences found between the reasons cited and actual predictive factors. Most notably, the retrospective factor of the total number of past edits was frequently talked about by voters but had little predictive impact on promotion success.

Our data is consistent with prior literature on contest theories of promotion which emphasize prior job performance as at least partially explanatory of promotion decisions (Turner, 1960). In our study, several metrics of a user’s prior experience were significantly related to promotion success. Importantly, we also introduce a distinction between prospective factors of evaluation which focus on a person’s likely success in their new role, and retrospective factors which emphasize past achievement. This distinction mirrors the dual role of promotion in assigning people to roles they are best able to contribute and as a source of incentives and rewards (Milgrom & Roberts, 1992). The greater importance of prospective factors in both the qualitative and quantitative analyses suggest that Wikipedia uses promotion as an assignment and selection mechanism for increased user privileges rather than as a reward mechanism for past achievement.

These results are also consistent with and extend theories of peer production as an efficient model for matching users with tasks to which they can maximally contribute (Benkler, 2002). Extant theories have focused on matching users with production tasks; e.g., users who are experts or have strong interests in a subject domain will be more likely to contribute to that domain. However, our data suggest that peer production provides a model for matching users with managerial tasks as well: candidates considered more likely to maximally contribute as
future administrators had increased promotion success. Peer production may thus be an efficient paradigm not only for matching users with production work but also for managerial work and promotion.

These findings also contribute to the literature on policy capture, in which several studies have found moderate differences between subjective estimates of policy importance and objective captured policy (Zedeck & Kafry, 1977; Hobson, Mendel, & Gibson, 1981; Stumpf & London, 1981). However, most extant studies of policy capture have employed artificial rating situations which some researchers argue are very different from actual settings and make generalization to real-world situations difficult (Hobson & Gibson, 1983). Here we demonstrate the benefits that archival online communities such as Wikipedia present as a way to study policy capture in actual promotion settings. Furthermore, we extend the study of policy capture to online communities, which are growing in real-world importance; as of January 2007, 6 of the top 10 visited internet sites involved online communities (Alexa, 2008).

Although this research has shown that judges pay attention to candidates’ job-relevant behavior and especially behavior that suggests the candidate will be a good manager and not just a good worker, it is silent about whether other factors identified in the organizational literature such as social networks, irrelevant attributes, or strategic self-presentation (Ng et al. 2005). Indeed, recent evidence that Wikipedia administrators use a secret mailing list to coordinate their actions toward others suggest that sponsorship may also play a role in promotion (Metz, 2007). Future research in Wikipedia using techniques like those in the current paper can be used to test theories in organizational behavior about criteria for promotion.

Another important limitation of the current model is that it does not take the quality of contribution into account. We plan to improve the model by examining measures of length,
persistence, and page views of edits, which are already being used in more processor intensive models of existing administrator behavior and impact of edits.

Additional research is also needed to determine whether the prospective criteria used by RfA voters to select administrators actually leads to better administrator performance. Our data currently only describe candidates who become administrators, not necessarily those who become good administrators. An important next step is to determine if administrators change their behavior after the RfA and to measure the quality of that administrator's future behavior.

**Conclusion**

In both a qualitative study of promotion explanation and a quantitative study of promotion success, people weighted prospective factors (predicted performance in a new role) more heavily than retrospective factors (previous role performance). This suggests that promotion in Wikipedia is primarily used as a device to assign people to roles in which they can maximally benefit the organization, rather than as an incentive or reward for past achievement. However, further research is needed to determine the role of other factors in administrator promotion and to link promotion decision criteria to post-promotion performance.
References


Figure 1: Criteria Used in RfA Voting

- Need for Admin Tools
- Vandalism
- Edit Diversity
- Work in Deletion Voting
- Knowledge of Policy
- Interaction with Others
- Trustworthy
- Neutral Point of View
- Conflict Management
- Community Involvement
- Commitment to Wikipedia
- Number of Edits
- Time in Wikipedia
- Article Quality

Positive Comment
Negative Comment

Number of RfA's that Cite Criteria
Table 1. Prospective and restrospective criteria as predictors of promotion to administrator status in Wikipedia.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>∆ Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prospective Task Criteria</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edit diversity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of sections edited</td>
<td>11</td>
<td>3.7</td>
<td>8.6% ***</td>
</tr>
<tr>
<td>Need for admin tools</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Admin attention/noticeboard edits</td>
<td>18.7</td>
<td>57.9</td>
<td>-2.6% †</td>
</tr>
<tr>
<td>Vandalism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Revert” in comments</td>
<td>257.6</td>
<td>563.2</td>
<td>3.1% †</td>
</tr>
<tr>
<td>Vandal-fighting (AIV) edits</td>
<td>26.5</td>
<td>108.7</td>
<td>-1.2% ns</td>
</tr>
<tr>
<td>Requests for protection</td>
<td>3.7</td>
<td>12.2</td>
<td>-1.6% ns</td>
</tr>
<tr>
<td>Work in deletion voting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X for deletion/review edits</td>
<td>504.5</td>
<td>1027.3</td>
<td>1.2% ns</td>
</tr>
<tr>
<td>Votes for deletion</td>
<td>3.4</td>
<td>15.1</td>
<td>1.1% ns</td>
</tr>
<tr>
<td><strong>Prospective Personal Criteria</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge of policy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wikipedia policy edits</td>
<td>433.8</td>
<td>913.8</td>
<td>7.4% *</td>
</tr>
<tr>
<td>Interaction with others</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Article talk edits</td>
<td>415.2</td>
<td>775.4</td>
<td>2.5% ns</td>
</tr>
<tr>
<td>User talk edits</td>
<td>786.6</td>
<td>1169.9</td>
<td>-5.9% **</td>
</tr>
<tr>
<td>Wikipedia talk edits</td>
<td>87.6</td>
<td>179.5</td>
<td>0.8% ns</td>
</tr>
<tr>
<td>“Please” in comments</td>
<td>31.7</td>
<td>83.8</td>
<td>-1.9% ns</td>
</tr>
<tr>
<td>“Thanks” in comments</td>
<td>21.8</td>
<td>39.3</td>
<td>7.6% ***</td>
</tr>
<tr>
<td>Neutral point of view</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“POV” in comments</td>
<td>26.7</td>
<td>46.9</td>
<td>3.5% *</td>
</tr>
<tr>
<td>Conflict management</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arb/mediation/wikiquette edits</td>
<td>9.8</td>
<td>47.1</td>
<td>-4.8% ***</td>
</tr>
<tr>
<td>Community involvement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other RfAs</td>
<td>27.4</td>
<td>58.9</td>
<td>0.9% ns</td>
</tr>
<tr>
<td>Village pump</td>
<td>9.5</td>
<td>34.6</td>
<td>-1.1% ns</td>
</tr>
<tr>
<td>Newcomer pump</td>
<td>76.9</td>
<td>321.1</td>
<td>-0.8% ns</td>
</tr>
<tr>
<td><strong>Retrospective Criteria</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of edits</td>
<td>Article edits</td>
<td>2611.</td>
<td>3804.3</td>
</tr>
<tr>
<td>WikiProject edits</td>
<td>144</td>
<td>569.1</td>
<td>-1.1% ns</td>
</tr>
<tr>
<td>Time in Wikipedia</td>
<td>Months since first edit</td>
<td>9.5</td>
<td>7.6</td>
</tr>
</tbody>
</table>

N=1502 *** p < .001 ** p < .01 * p < .05 † p < 0.1