
Aaron Halfaker
Aniket Kittur (CMU)
Robert Kraut (CMU)
John Riedl
Presentation Outline

1) Formal peer review and Wikipedia
2) Analysis of review in Wikipedia
3) Results
   • Quality
   • Experience
   • Ownership
4) Conclusions
Presentation Outline

1) Formal peer review and Wikipedia

2) **Analysis of review in Wikipedia**

3) Results
   - Quality
   - Experience
   - Ownership

4) Conclusions
Presentation Outline

1) Formal peer review and Wikipedia
2) Analysis of review in Wikipedia
3) Results
   • Quality
   • Experience
   • Ownership
4) Conclusions
Presentation Outline

1) Formal peer review and Wikipedia
2) Analysis of review in Wikipedia
3) Results
   • Quality
   • Experience
   • Ownership
4) Conclusions
Peer review

Peer review (also known as refereeing) is the process of subjecting an author's scholarly work, research, or ideas to the scrutiny of others who are experts in the same field. Peer review requires a community of experts in a given (and often narrowly defined) field, who are qualified and able to perform impartial review. Impartial review, especially of work in less narrowly defined or inter-disciplinary fields, may be difficult to accomplish; and the significance (good or bad) of an idea may never be widely appreciated among its contemporaries. Although generally considered essential to academic quality, peer review has been criticized as ineffective, slow, and misunderstood (see anonymous peer review and open peer review).

Pragmatically, peer review refers to the work done during the screening of submitted manuscripts and funding applications. This process encourages authors to meet the accepted standards of their discipline and prevents the dissemination of irrelevant findings, unwarranted claims, unacceptable interpretations, and personal views. Publications that have not undergone peer review are likely to be regarded with suspicion by scholars and professionals.
Peer review is the process of subjecting an author's work, research or ideas to the scrutiny of their peers in the same field.
Peer review (also known as refereeing) is the process of subjecting an author's scholarly work, research, or ideas to the scrutiny of others who are experts in the same field. Peer review requires a community of experts to come together to evaluate the validity of any new conclusions or interpretations. This process encourages authors to meet the accepted standards of their discipline and prevents the dissemination of irrelevant findings, unwarranted claims, [...]
Peer Review Systems

- Peer reviewed conferences (like WikiSym)
- NSF grant panels
- Academic journals
- Art competitions
Peer Review Systems

- Peer reviewed conferences (like WikiSym)
- NSF grant panels
- Academic journals
- Art competitions
- Wikipedia?
Traditional Peer Review

1. **Produce complete work**
2. **Submit for review**
3. **Peer approval?**
   - **Yes** → **Publication**
   - **No** → **Rejection**
Traditional Peer Review

Produce complete work

Submit for review

Peer approval?
Yes
Publication

No
Rejection
Traditional Peer Review

Produce complete work

Submit for review

Peer approval?

Yes → Publication

No → Rejection
Traditional Peer Review

1. Produce complete work
2. Submit for review
3. Peer approval?
   - Yes: Publication
   - No: Rejection
Wikipedia Peer Review

1. Edit article
2. Submit change
3. Publication
4. Peer approval?
   - Yes: The change persists
   - No: Revert
Wikipedia Peer Review

Edit article

Submit change

Publication

Peer approval?

Yes: The change persists

No: Revert
Wikipedia Peer Review

1. Edit article
2. Submit change
3. Publication
   - Peer approval?
     - Yes: The change persists
     - No: Revert
Wikipedia Peer Review

1. Edit article
2. Submit change
3. Publication
4. Peer approval?
   - Yes: The change persists
   - No: Revert
Wikipedia Peer Review

1. Edit article
2. Submit change
3. Publication
4. Peer approval?
   - Yes: The change persists
   - No: Revert
Analysis of Traditional Peer Review

• Expert Evaluation
  • Goal: What is the quality of a review?
    – How many rejections should have been accepted?

• Statistical Model
  • Goal: What predicts the outcome of a review?
    – Is past acceptance highly correlated with future acceptance?
    – Are submissions that are critical of reviewers' previous work correlated with rejection?
Our approach: Statistical

- Can't determine the quality of a review
- Resilient to researcher bias.
- Allows us to compare the effects of various metrics in predicting outcome
  - Quality
  - Experience
  - Ownership
- Answers: How do quality, experience and ownership effect the outcome of review—being reverted?
Statistical Model

• Sample
  • Observation = revision
  • n = 1.4 million
  • Boolean outcome: was the revision eventually reverted?
Hypothetical “Ideal” Model

Experience

Ownership

Quality

Probability of revert
Hypothetical “Ideal” Model

Experience

Ownership

Quality

Probability of revert
Hypothetical “Ideal” Model

Experience

Ownership

Quality

Probability of revert
Wikipedia:Ownership of articles

From Wikipedia, the free encyclopedia
(Redirected from Wikipedia:Own)

This page documents an English Wikipedia policy, a widely accepted standard that all editors should normally follow. Changes made to it should reflect consensus.

This page in a nutshell: You do not own articles (nor templates and other features of Wikipedia). If you create or edit an article, know that others will edit it, and within reason you should not prevent them from doing so.

This page is about control over an article's text. For the ownership of copyright in an article, see Wikipedia:Copyrights.

All Wikipedia content is edited collaboratively. Wikipedia contributors are editors, not authors, and no one, no matter how skilled, has the right to act as if they are the owner of a particular article.

## Contents

1. Overview
   1.1 Use of signatures
2. Types of ownership
   2.1 Primary editors
   2.2 Multiple editors
3. Resolving ownership issues
4. Notes
5. See also

Wikipedia policy

Behavioral standards
- Bots
- Civility
- Editing policy
- No legal threats
- No personal attacks

Ownership of articles
- Sock puppetry
- Three-revert rule
- User accounts
- Vandalism
[...] no one, no matter how skilled, has the right to act as if they are the owner of a particular article.
Hypothetical “Ideal” Model

- Experience
- Ownership

Quality

Probability of revert

0

−

−
Hypothetical “Ideal” Model

Experience

Ownership

Quality
- Removed word persistence
- Recent edit quality
- Recent success rate

Probability of revert

University of Minnesota
A Jury of Your Peers
Quality = Word Persistence

Assumption: Words that last are part of a high quality edit.

1 Steve: blue apples are yummy
2 Chris: apples are yummy
3 Paul: apples are certainly yummy
4 Robin: apples are certainly most yummy
5 Phil: apples are nutritious
6 Dave: apples are nutritious for realz
Quality = Word Persistence

Application

- **Removed Word Persistence:** Edits that remove established words are removing high quality content and are likely to be low quality changes.

- **Recent Edit Quality:** Editors who have recently contributed content that lasts several revisions are high quality editors.

Justification

- Significant correlation between high word persistence editors and increased in the Wikipedia 1.0 Assessment rating.
Hypothetical “Ideal” Model

Experience

Ownership

Quality
- Removed word persistence
- Recent edit quality
- Recent success rate

Probability of revert

0

-
Hypothetical “Ideal” Model

**Experience**
- Tenure
- Previous edits
- Citations to policy

**Ownership**

**Quality**
- Removed word persistence
- Recent edit quality
- Recent success rate

**Probability of revert**
Hypothetical “Ideal” Model

Experience

- Tenure
- Edits
- Citations to policy

Ownership

- Stepping on toes

Quality

- Removed word persistence
- Recent edit quality
- Recent success rate

Probability of revert

0
Stepping on Toes

“Removing words added by editors when they are likely to notice.”

1 Steve: **blue apples are yummy**

2 Chris: **apples are yummy** (rm: Steve)

3 Paul: **apples are certainly yummy**

4 Robin: **apples are certainly most yummy**

5 Phil: **apples are nutritious** (rm: Steve, Robin, Paul)

6 Dave: **apples are nutritious for realz**
Hypothetical “Ideal” Model

Experience
- Tenure
- Edits
- Citations to policy

Ownership
- Stepping on toes

Quality
- Removed word persistence
- Recent edit quality
- Recent success rate

Probability of revert
Results

- Quality
- Experience
- Ownership
Quality

- Removed word persistence
- Recent edit quality
- Recent success rate
Recent quality is a powerful predictor.

- Removed word persistence
- Recent edit quality
- Recent success rate
Recent quality is a powerful predictor.

- Removed word persistence
- Recent edit quality
- Recent success rate
Experience
Editors don't exhibit a learning effect.
Editors don't exhibit a learning effect.

**Tenure:** time since editor's first edit

**Lifetime:** time an editor will continue to edit

[Graph showing revisions over time by all editors and editors that last 52 weeks.]
Editors don't exhibit a learning effect.

Tenure: time since editor's first edit
Lifetime: time an editor will continue to edit

Tenure: \( p = 0.37 \)
Lifetime: \( p < 0.0001 \)

The predictive power of lifetime consumes tenure.
Ownership
Ownership has a powerful effect over which contributions are rejected.
Ownership has a powerful effect over which contributions are rejected.

- Very powerful predictor
Ownership has a powerful effect over which contributions are rejected.

- Very powerful predictor
- Independent
  - Experience
  - Quality of work
  - Recent success rate
Ownership has a powerful effect over which contributions are rejected.

- Very powerful predictor
- Independent
  - Experience
  - Quality of work
  - Recent success rate
Ownership has a powerful effect over which contributions are rejected.

- Very powerful predictor
- Independent
  - Experience
  - Quality of work
  - Recent success rate

![Graph showing the probability of being reverted against active editors with words removed in revision (log scaled).]
Ownership has a powerful effect over which contributions are rejected.

- Very powerful predictor
- Independent
  - Experience
  - Quality of work
  - Recent success rate
Ownership has a powerful effect over which contributions are rejected.

- Very powerful predictor
- Independent
  - Experience
  - Quality of work
  - Recent success rate

If you step on other editors' toes, you're likely to be reverted.
Resulting Model

Experience
- Tenure
- Previous edits
- Citations to policy

Ownership
- Stepping on toes

Quality
- Removed word persistence
- Recent edit quality
- Recent success rate

Probability of revert
Resulting Model

Quality
- As hypothesized
- Neg. correlated with revert prob.

Experience
- Tenure
- Previous edits
- Citations to policy

Ownership
- Stepping on toes

Quality
- Removed word persistence
- Recent edit quality
- Recent success rate

Probability of revert
Resulting Model

Experience

- Tenure
- Previous edits
- Citations to policy

Quality

- As hypothesized
- Neg. correlated with revert prob.

Experience

- No effect after controlling for lifetime/total edits

Ownership

- Stepping on toes

Probability of revert

- Removed word persistence
- Recent edit quality
- Recent success rate
Resulting Model

Quality
- As hypothesized
- Neg. correlated with revert prob.

Experience
- No effect after controlling for lifetime/total edits

Experience
- Tenure
- Previous edits
- Citations to policy

Ownership
- Stepping on toes

Quality
- Removed word persistence
- Recent edit quality
- Recent success rate

Probability of revert

University of Minnesota
A Jury of Your Peers
Resulting Model

Quality
- As hypothesized
- Neg. correlated with revert prob.

Experience
- No effect after controlling for lifetime/total edits

Ownership

Experience
- Tenure
- Previous edits
- Citations to policy

Ownership
- Stepping on toes

Quality
- Removed word persistence
- Recent edit quality
- Recent success rate

Probability of revert
Resulting Model

- **Quality**
  - As hypothesized
  - Neg. correlated with revert prob.

- **Experience**
  - No effect after controlling for lifetime/total edits

- **Ownership**
  - Powerful and independently correlated with revert prob.

**Experience**
- Tenure
- Previous edits
- Citations to policy

**Ownership**
- Stepping on toes

**Quality**
- Removed word persistence
- Recent edit quality
- Recent success rate

**Probability of revert**

0

+ 
-
Contributions/Summary

We constructed a logistic regression model to estimate the effects of various editor and edit characteristics on rejection in Wikipedia.

- Verified word persistence as a robust metric for perceived quality.
- Editors do not show a learning effect.
- Ownership of content has a powerful effect.
Acknowledgments

- My co-authors
  - Niki Kittur - nkittur@cs.cmu.edu
  - Robert Kraut - robert.kraut@cs.cmu.edu
  - John Riedl - riedl@cs.umn.edu
- The Wikipedia group in the GroupLens lab
- This work has been financially supported by the National Science Foundation: IIS 05-34420

Aaron Halfaker
halfak@cs.umn.edu
Table 1: Two logistic regression coefficients and p-values. “All applicable revisions” covers all of the revisions in the sample. “Revisions by old editors” covers a revisions that were made by editors after they were 90 days old. For the discussion, statistical significance corresponds to $\alpha = 0.01$.

<table>
<thead>
<tr>
<th></th>
<th>All Revisions</th>
<th></th>
<th>Revision by old editors</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>$P(&gt;</td>
<td>z</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-3.512</td>
<td>.008</td>
<td>&lt; .0001</td>
<td>-3.640</td>
</tr>
<tr>
<td>Total days an editor will be active</td>
<td>-0.098</td>
<td>.010</td>
<td>&lt; .0001</td>
<td>-0.062</td>
</tr>
<tr>
<td>(total days)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recent quality (log PWRpW of last 20</td>
<td>-0.183</td>
<td>.007</td>
<td>&lt; .0001</td>
<td>-0.161</td>
</tr>
<tr>
<td>editing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Days since an editor began editing</td>
<td>-0.008</td>
<td>.009</td>
<td>0.3742</td>
<td>0.029</td>
</tr>
<tr>
<td>(current tenure)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous additions to talk pages</td>
<td>-0.018</td>
<td>.012</td>
<td>0.1311</td>
<td>-0.024</td>
</tr>
<tr>
<td>citing policy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience via completed sessions</td>
<td>-0.252</td>
<td>.010</td>
<td>&lt; .0001</td>
<td>-0.159</td>
</tr>
<tr>
<td>(%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion edits recently reverted</td>
<td>0.318</td>
<td>.004</td>
<td>&lt; .0001</td>
<td>0.276</td>
</tr>
<tr>
<td>for non-vandalism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion edits recently reverted</td>
<td>0.217</td>
<td>.003</td>
<td>&lt; .0001</td>
<td>0.128</td>
</tr>
<tr>
<td>for vandalism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edits reverting other editors</td>
<td>0.053</td>
<td>.004</td>
<td>&lt; .0001</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active editors w/words removed</td>
<td>0.231</td>
<td>.004</td>
<td>&lt; .0001</td>
<td>0.249</td>
</tr>
<tr>
<td>(log active editors removed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Persistence of removed words (log</td>
<td>0.164</td>
<td>.006</td>
<td>&lt; .0001</td>
<td>0.128</td>
</tr>
<tr>
<td>PWRpW of removed words)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of words added by edit</td>
<td>0.024</td>
<td>.004</td>
<td>&lt; .0001</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of words removed by edit</td>
<td>0.006</td>
<td>.002</td>
<td>0.008</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction between recent quality</td>
<td>0.016</td>
<td>.005</td>
<td>0.002</td>
<td>0.019</td>
</tr>
<tr>
<td>and persistence of removed words</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Correlation table of explanatory variables.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Total days an editor will be</td>
<td>1.0</td>
<td>.13</td>
<td>.58</td>
<td>.00</td>
<td>.40</td>
<td>-.13</td>
<td>-.12</td>
<td>.09</td>
<td>.02</td>
<td>-.06</td>
<td>-.01</td>
<td>.00</td>
</tr>
<tr>
<td>active</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Recent quality</td>
<td>.13</td>
<td>1.0</td>
<td>-.07</td>
<td>-.03</td>
<td>-.22</td>
<td>-.11</td>
<td>-.09</td>
<td>-.17</td>
<td>-.13</td>
<td>.10</td>
<td>-.08</td>
<td>-.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Days since an editor began editing</td>
<td>.58</td>
<td>.07</td>
<td>1.0</td>
<td>.02</td>
<td>.62</td>
<td>-.12</td>
<td>-.16</td>
<td>.19</td>
<td>.06</td>
<td>.03</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Previous additions to talk pages</td>
<td>.00</td>
<td>-.03</td>
<td>.02</td>
<td>1.0</td>
<td>.04</td>
<td>.00</td>
<td>.00</td>
<td>.03</td>
<td>.00</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>citing policy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Experience via completed</td>
<td>.40</td>
<td>-.22</td>
<td>.62</td>
<td>.04</td>
<td>1.0</td>
<td>-.15</td>
<td>-.14</td>
<td>.32</td>
<td>.04</td>
<td>-.05</td>
<td>.02</td>
<td>.00</td>
</tr>
<tr>
<td>sessions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Edits recently reverted for</td>
<td>-.13</td>
<td>-.11</td>
<td>-.12</td>
<td>.00</td>
<td>-.15</td>
<td>1.0</td>
<td>.09</td>
<td>.03</td>
<td>.09</td>
<td>.10</td>
<td>.03</td>
<td>.02</td>
</tr>
<tr>
<td>non-vandalism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Edits recently reverted for</td>
<td>-.12</td>
<td>-.09</td>
<td>-.16</td>
<td>.00</td>
<td>-.14</td>
<td>.09</td>
<td>1.0</td>
<td>-.01</td>
<td>.02</td>
<td>.05</td>
<td>.02</td>
<td>.01</td>
</tr>
<tr>
<td>vandalism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Edits reverting other editors</td>
<td>.09</td>
<td>-.17</td>
<td>.19</td>
<td>.03</td>
<td>.32</td>
<td>.03</td>
<td>-.01</td>
<td>1.0</td>
<td>.13</td>
<td>.03</td>
<td>.02</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Active editors w/words removed</td>
<td>.02</td>
<td>-.13</td>
<td>.06</td>
<td>.00</td>
<td>.04</td>
<td>.09</td>
<td>.02</td>
<td>.13</td>
<td>1.0</td>
<td>.13</td>
<td>.07</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Persistence of removed words</td>
<td>-.06</td>
<td>.10</td>
<td>.03</td>
<td>.01</td>
<td>-.05</td>
<td>.10</td>
<td>.05</td>
<td>.03</td>
<td>.13</td>
<td>1.0</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Number of words added by edit</td>
<td>-.01</td>
<td>-.08</td>
<td>.00</td>
<td>.00</td>
<td>.02</td>
<td>.03</td>
<td>.02</td>
<td>.02</td>
<td>.07</td>
<td>.00</td>
<td>1.0</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Number of words removed by edit</td>
<td>.00</td>
<td>-.02</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.02</td>
<td>.01</td>
<td>.01</td>
<td>.07</td>
<td>.01</td>
<td>.03</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Recent quality is a powerful predictor.

- Removed word persistence
- Recent edit quality
- Recent success rate

![Graph showing probability of being reverted (log scaled) versus average revisions of words removed (log scaled).]