

## Quantifying the work of copy editors

**Yateendra Joshi**

*Cactus Communications, 510 Shalimar Morya Park, off Link Road, Andheri (West), Mumbai 400 053, India*  
*yateendra.joshi@gmail.com*

**Abstract** Copy editors typically work on word-processor files, which track changes made to the text. They also query the authors through comments in the text. At present there is no automated way to quantify the copy editing, and the amount of work is typically assessed by word count of the original document or the time spent by the copy editor. These are poor substitutes, and do not really measure what has been done to the manuscript. This article proposes some approaches to quantifying a range of corrections made by copy editors.

**Keywords** Copy editing; editorial changes; manuscript editing

### Introduction

Nearly all journal submissions require editing, although the extent of the editing varies across manuscripts and journals. Van Buren and Buehler<sup>1</sup> elaborate nine levels of edit, with substantive editing representing the highest level, which the University of Chicago Press<sup>2</sup> describes as dealing with ‘the organization and presentation of content.’ Copy editing, on the other hand, is concerned with the exact words in which that subject matter is couched, including spelling, grammar, punctuation, style, and usage. Most editing services typically copy-edit the manuscripts, and publishers encourage authors – especially non-native English speakers – to have their manuscripts copy-edited and even refer to agencies that offer such services.

For authors, a simple criterion to judge the value of copy editing is the outcome of their submission: if the paper is accepted for publication or if the reviewers make no comment on language or grammar, the copy editing is considered worthwhile. It is important not only to judge whether the job has been done adequately but also to know how much work was put into copy editing. Just as the individual *h* index serves as a tool for assessing researchers’ performance without looking into the contents of their papers, those who supervise copy editors need an approach to measure the amount of editing without examining the copy-edited manuscript line by line.

The present article discusses some approaches to quantifying a range of corrections made by copy editors.

A copy-edited manuscript is usually dotted with many small changes and queries to the author but it will seldom show text shifted even within a paragraph. Substantive editing, on the other hand, will be visible by directions to move blocks of text, large-scale deletions and additions, and queries to the author about the logic and organization of text. Schultz<sup>3</sup> provides a simple diagram, referred to as the writing/editing funnel, in which organization and paragraphs represent the top of the funnel and words, punctuation, grammar, etc. represent the narrow end—copy editing focuses on the narrow end.

### Indirect ways of quantifying copy editing

At present, copy editing services charge their customers based on the number of words in the manuscript, time spent on editing, or a flat fee quoted after a preliminary assessment of the manuscript. Word count has the merit that the customers know exactly how much the service is going to cost them. Some editors or editing services charge on the basis of number of pages. However, a page is generally specified as 250 words, which means the charges are essentially based on the word count.

Some editors charge an hourly fee and keep track of the time spent on each manuscript. Although this approach ensures that the copy editor is adequately compensated for manuscripts that require heavy copy editing, the customers cannot know the exact fees.

The level of editing is also specified, ranging from light to heavy editing, and the service is charged for accordingly. None of these methods, however, can take into account the extent of changes made by the copy editor. Fewer changes do not necessarily mean inadequate editing though: a manuscript may be well written; its authors may have scrupulously followed the journal’s instructions to authors; and stylistic inconsistencies may have been eliminated by using a software package. It is also possible that heavy editing has added little value, most of the changes being the editor’s pet peeves or stylistic preferences.

### More direct ways of quantifying copy editing

Whereas the indirect ways mentioned above attempt to predict the amount of editing a given manuscript is likely to require, the more direct ways involve comparing the original and the copy edited versions by means that are largely automated; it is not necessary to actually read the copy edited version.

#### Word count

A competent copy editor eliminates verbiage; it follows, therefore, that the copy edited version is shorter than the original, and it is possible to quantify this: open a copy of the revised version; accept all changes; remove embedded comments or queries; and compare the word count of this copy with that of the original.

#### Readability statistics

Microsoft Word, for example, can calculate the readability statistics of a given file, and comparing these statistics for the original and the edited versions can show whether the editing has contributed to making the text more readable.

#### Vocabulary

Competent copy editors, particularly those familiar with the subject of the manuscript, often use the right words or technical terms where authors may have used less precise

terms. A software package that can compile a concordance is useful here and can show at a glance whether the edited version has introduced words not found in the original.

### Categorizing the changes made to a manuscript

Ideally we need a method that can compile a detailed profile of a copy-editing job just as chemical analysis of a sample of liquid, for example, supplies its profile: the elements that make up the compound and the quantity of each, the pH, turbidity, and so on.

Wates and Campbell<sup>4</sup> compared the original, that is as submitted, and the published versions of 189 papers from 23 journals and found that citations and references was the single largest category of changes (42.7%) made to the manuscripts. Boettger<sup>5</sup> categorized the errors that copy editors were required to spot in 41 such tests used by different agencies. Of the 20 categories, spellings made up the largest category (found in about 75% of the tests), followed by inappropriate or missing capitalization (about 65%) and missing comma with a non-restrictive element (65%). Such errors are typically the kind of errors that copy editing is expected to eliminate.

One of the earliest studies on categorizing the changes made by copy editors is that by Portugal and Forscher,<sup>6</sup> who grouped the defects that copy editors corrected into four broad areas, namely nomenclature, spelling, punctuation, and construction (subject-verb agreement, tense, etc.) and found “remarkable degree of uniformity among the editors who [had] similar training but different levels of editing experience.” A longer list of categories of the changes made to manuscripts is shown in Table 1.

### Quantifying copy editing

It is only by quantifying at least some items in the list given in Table 1 that we can approach the task of quantifying copy editing. To my mind, macros offer the key: it should be possible to automate the counting of one or more categories of the changes listed here. For example, multiple changes within a sequence of, say, 20 consecutive words probably imply a re-write; if the bulk of the changes are only those involving spelling, capitalization, and formatting, the job can be categorized as a light edit; if the changes predominantly involve prepositions and articles, the author is probably a non-native speaker.

I believe that such a task should not be unduly difficult for those who are well versed in macros and for the developers of style and grammar checkers. In fact, the process of developing such macros or similar utilities may even contribute to refining style and grammar checkers. Secondly, just as dictionaries are now increasingly corpus-based, data from large-scale analyses of errors fixed by copy editors can contribute to corpus-based style manuals. Lastly, just as the h index, citation counts, and other similar measures offer the tools to evaluate individual researchers' performance, quantifying copy editing can pave the way for correctly crediting copy editors' efforts.

**Note** The essay is based on a presentation by the author at the 11th International Conference of EASE, Tallinn, June 2012.

### References

- 1 Van Buren R, Buehler MF. *The Levels of Edit*. 2nd ed. Pasadena, California: Jet Propulsion Laboratory;1980.
- 2 University of Chicago Press. *The Chicago Manual of Style*. 16th ed. Chicago: University of Chicago Press. 2010.
- 3 Schultz D. *Eloquent Science: a practical guide to becoming a better writer, speaker, and atmospheric scientist*. Boston, Massachusetts: American Meteorological Society. 2009.
- 4 Wates E, Campbell R. Author's version vs. publisher's version: an analysis of the copy-editing function. *Learned Publishing* 2007;20(2):121–129. doi: 10.1087/174148507X185090
- 5 Boettger RK. Examining error in the technical communication editing test. Proceedings of the 58th Annual Technical Communication Summit, Sacramento, California;2011:142–147.
- 6 Portugal FH, Forscher BK. Can editing be quantitated? In: Balaban M (ed.) *Scientific Information Transfer: the editor's role*. Dordrecht: Reidel, 1978, 265–270. [Proceedings of the First International Conference of Scientific Editors, 24–29 April 1977, Jerusalem]

**Table 1 Categories of changes made to manuscripts in copy editing**

**Note** By adding three columns, namely deletion, addition, and substitution, the table becomes a matrix, which can be useful in profiling any copy editing job.

#### Character

Spelling  
Capitalization  
Spacing  
Symbols  
Fonts  
Consistency  
Punctuation

#### Word

Homophones  
Sense-related  
Prepositions  
Articles  
Tenses  
Tautology

#### Phrase

Idiomatic usage

#### Sentence

Re-writing  
Deleting superfluous words  
Readability  
Active/passive  
De-nominalization

#### Citations and references

Format (capitalization, italics, boldface, etc.)  
Sequence of elements  
Punctuation  
Omissions  
Extraneous entries  
Mismatches

#### Housekeeping

Tables  
Figures

#### Formatting