

## Essays

### How can we evaluate the quality of abstracts?

**Sylwia B Ufnalska**

Freelance translator and editor, Poland; sylwia.ufnalska@gmail.com

**James Hartley**

Research Professor, School of Psychology, Keele University, UK; j.hartley@psy.keele.ac.uk

	Abstract 1	Abstract 2	Abstract 3
How do you judge an abstract to be good or bad? Do you just feel it, or do you have more systematic ways of doing this? We will discuss three methods for evaluating abstracts, indicating their pros and cons. First we will ask you to rate three abstracts and to compare your ratings with those of other judges. Then we shall contrast such rating scales with two other methods, namely checklists and readability measures. Our aim is to allow you to compare these three methods: choose one or two or three – whichever work best for you – and use them when there is a need.			
Understandability	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Grammar	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Spelling	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Structure	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Selection of information	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Brevity	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Suitability for international readership	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
<b>Rating scales</b> Please read carefully each of the three abstracts below and rate them in respect of the characteristics listed on the scale: 1 = very poor, 2 = poor, 3 = average, 4 = very good, 5 = excellent.			

*Abstract 1. Evidence that mercury from silver dental fillings may be an etiological factor in multiple sclerosis*

This paper investigates the hypothesis that mercury from silver dental fillings (amalgam) may be related to multiple sclerosis (MS). It compares blood findings between MS subjects who had their amalgams removed to MS subjects with amalgams. MS subjects with amalgams were found to have significantly lower levels of red blood cells, hemoglobin and hematocrit compared to MS subjects with amalgam removal. Thyroxine levels were also significantly lower in the MS amalgam group and they had significantly lower levels of total T Lymphocytes and T-8 (CD8) suppressor cells. The MS amalgam group had significantly higher blood urea nitrogen and lower serum IgG. Hair mercury was significantly higher in the MS subjects compared to the non-MS control group. A health questionnaire found that MS subjects with amalgams had significantly more (33.7%) exacerbations during the past 12 months compared to the MS volunteers with amalgam removal. The paper also examines epidemiological correlations between dental caries and MS, as well as how mercury could be causing pathological and physiological changes found in multiple sclerosis.

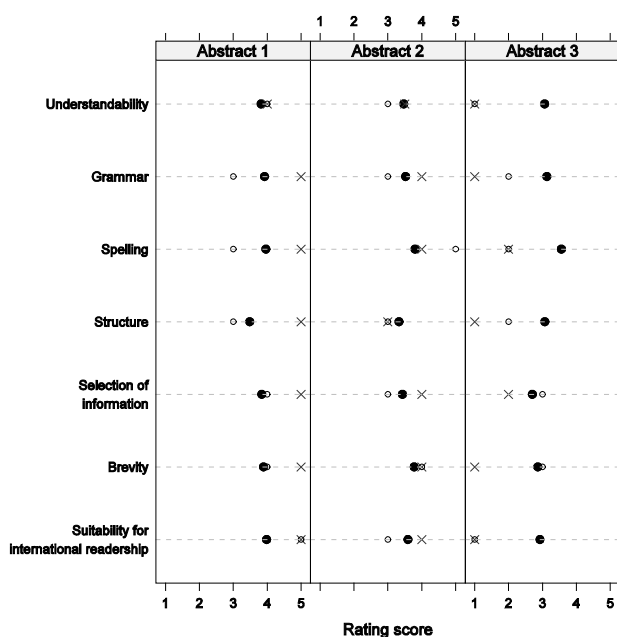
*Abstract 2. Accumulation and use of nitrogen and phosphorus following fertilization in two alpine tundra communities*

To determine whether there are differences in the relative capacity of communities to accumulate nutrients and translate nutrient uptake into growth, N and P standing crops and use efficiencies were measured following fertilization in two alpine tundra communities. In general, differences in nutrient dynamics between the communities corresponded with the type and degree of nutrient limitation of production. The N-limited dry meadow had greater increases in aboveground N standing crop and tissue N concentration in response to N fertilization, higher N-use efficiency, and higher N resorption than the wet meadow. Conversely the N-P co-limited wet meadow had a greater P accumulation response to P fertilization and higher P-use efficiency than the dry meadow. Differences in the response to the treatments and in nutrient use efficiencies were mediated largely by individual plant growth forms. Although there was a substantial amount of luxury consumption of N and P, there was evidence of co-regulated uptake relative to the availability of these nutrients in the soil.

*Abstract 3. Relation of berries crop to measure features of overground parts of blueberry (Vaccinium myrtillus L.) defined for huge and small harvest*

Examinations were conducted according to cropping of blueberry in relation to features of its overground organs. Results based on examination showed that relation between berry crop of blueberry and chosen group of measure features of overground twigs in years of huge harvest is higher than similar relation in small harvest years. Besides it was concluded that number of twigs (observations) in that does not guarantee the increase of level of researched relation.

Which one of these abstracts is best in your opinion? Clearly there is no one way of assessing their quality. However, with rating scales, you can compare your results with those of others. In this paper the open circles and the crosses in Figure 1 show the ratings of two highly experienced scientists (ecologists), native speakers of English. The solid circles show the mean ratings of 33 postgraduate students (all non-native English speakers) on an international Summer Program in Environmental Sciences and Policy organized by the Budapest College of the Central European University.<sup>1</sup>



**Figure 1** Comparison of the evaluation of the sample abstracts by highly experienced native speakers of English (x = Expert 1; o = Expert 2) and by an international group of postgraduate students (• mean value).

Abstracts 1 and 2 are published abstracts,<sup>2,3</sup> and Abstract 3 is an unpublished text. The data in Figure 1 show that Abstract 1 usually received very good scores, Abstract 2 did slightly less well, and Abstract 3 was generally regarded as average by students, but as poor by both experienced authors. Thus the students apparently evaluated the abstracts quite adequately, but seemed reluctant to use the highest and the lowest scores (perhaps they did not regard themselves as sufficiently qualified to judge). However, most of them have not noticed spelling mistakes in Abstract 3. (In fact, even one of the experts did not notice a spelling mistake in the title of Abstract 1.) Some remarkable differences can be noticed between the ratings made by the two native speakers, but still their overall rankings of the abstracts were similar to the mean ranking by students.

It is fairly common to use rating scales in this way to evaluate abstracts but it is not without difficulties. In this essay we contrast this method of evaluating abstracts with

two other methods – checklists and readability scores – to illustrate the strengths and weaknesses of different measures.

**Checklists**

Checklists require the reader to note the presence or absence of particular features of abstracts, but not to rate them for their quality. Figure 2 provides a brief checklist.<sup>4</sup> You might like to complete it with respect to the three abstracts above: please tick each box that applies for each abstract.

	Abstract 1	Abstract 2	Abstract 3
Background			
Aim			
Methods			
Results			
Conclusions			

**Figure 2** Short checklist to assess key features in abstracts.

Do you think all of the items on the checklist are adequately covered in the above abstracts? We judge, for example, that the aims, methods, and results are included in Abstract 1, while the conclusions are not clearly stated, but they are explicit in the title (a common procedure in medical journals). This suggests that abstracts should be evaluated jointly with the title. In Abstract 1, the background to the study (that led the investigators to want to carry it out) is not given. However, one may ask: does the background always need to be explained in the abstract? An educated reader knows that mercury is toxic and that multiple sclerosis is an important neurological disease, so the lack of background in Abstract 1 may seem justifiable. However, if an author seeks to attract the attention of a wider range of readers, such background information can be helpful.

Other investigators have provided different checklists for evaluating abstracts, and most contain many more items than the five basic ones listed above.<sup>4</sup> This is especially true in the medical field. Whatever the case, authors and editors should make sure that the abstract is not misleading and that it adequately reflects the content of the article. It is also important to check that all of the information found in the abstract is included in the main body of the article.

**Readability measures**

A different way of evaluating abstracts is to judge how readable they are, and thus how suitable they might be for a particular readership. One such measure, the Flesch Reading Ease score, is based on the length of sentences and the length of words in these sentences.<sup>5</sup> The scores range from 0 (very difficult) to 100 (very easy). A score in the mid-20s is typical of abstracts and scientific articles in general. This means that they are not easy to read. Indeed, abstracts are often more difficult to read than are Introductions, which are often more difficult to read than are Discussions.<sup>6</sup> In our examples, the Flesch score is 26.4 for Abstract 1, 19.6 for Abstract 2, and 34.4 for Abstract 3. This would suggest that the first abstract is more readable than the second, which

agrees with results of their rating by readers. However, the most readable would be the third one, although this abstract scored lowest in the ratings presented above. This disparity shows that the Flesch Reading Ease measure alone is not a reliable method of assessing the quality of abstracts. It can be used, however, to flag up excessively long words and sentences.

### Final remarks

We think that there is no ideal way of assessing the quality of abstracts: different methods have different strengths and weaknesses, and different readers will have different expertise in the subject matters of the papers that they are reading. Reader evaluations are possibly the most useful, although they can be substituted by a checklist combined with some form of readability measurement, if the spelling and grammar pose no problems.

The readability of abstracts can be enhanced or reduced by their typographic settings.<sup>7</sup> Unfortunately, abstracts are often printed in smaller typefaces than the ones used in the main text. For the sake of readers this practice should be discouraged because abstracts are read much more often than whole articles. Furthermore, abstracts are typically presented as a single block of text (as in the examples here). Separating the components of abstracts under subheadings (to create structured abstracts) can make them more readable.<sup>8</sup>

Linguistic data suggest that some problems with scientific abstracts written in English may be characteristic of authors who are non-native speakers of English, as expectations differ between countries about what is the best way to proceed.<sup>9, 10, 11</sup> Martin, for example, shows that many Spanish authors may omit the results in their abstracts.<sup>9</sup> For this reason, we suggest that guidebooks about scientific writing in English should be published in languages other than English. For example, such a guidebook has already been published in Polish.<sup>12</sup>

Swales and Feak provide useful suggestions on writing abstracts, especially for non-native speakers of English.<sup>13</sup> As a general rule, though, it is useful for authors to complete a brief checklist, to make sure that the abstract reflects the content of the article; and to ask one or two colleagues to rate the quality of the abstract before submitting the manuscript. Thanks to this, the abstract should be improved and the manuscript more likely to be accepted for publication.

We are very grateful to Elise Langdon-Neuner and Marcin Kozak for their encouragement and help. We also thank Stuart Handysides, Ed Hull, and an anonymous reviewer for some useful comments about an early version of this manuscript.

### References

- 1 Ufnalska S. Abstracts of research articles: readers' expectations and guidelines for authors. *European Science Editing* 2008;34(3):63–65.
- 2 Sibling RL, Kienholz E. Evidence that mercury from silver dental fillings may be an etiological factor in multiple sclerosis. *The Science of the Total Environment* 1994;142:191–205.
- 3 Bowman WD. Accumulation and use of nitrogen and phosphorus following fertilization in two alpine tundra communities. *Oikos* 1994;70:261–270.
- 4 Hartley J, Betts L. Common weaknesses in traditional abstracts in the social sciences. *Journal of the American Society for Information Science and Technology* 2009 (in press).
- 5 Flesch R. A new readability yardstick. *Journal of Applied Psychology* 1948;32: 221–233.
- 6 Hartley J, Pennebaker JW, Fox C. Abstracts, introductions and discussions: How far do they differ in style. *Scientometrics* 2003;57(3):389–398.
- 7 Hartley J. Typographic settings for structured abstracts. *Journal of Technical Writing and Communication* 2000;30(4):355–365.
- 8 Hartley J, Betts L. The effects of spacing and titles on judgments of the effectiveness of structured abstracts. *Journal of the American Society for Information Science and Technology* 2007;58(14):2335–2340.
- 9 Martin PM. A genre analysis of English and Spanish research paper abstracts in experimental social sciences. *English for Specific Purposes* 2003;22:225–243.
- 10 Sauperl S, Klasinc J, Luzar S. Components of abstracts: Logical structure of scholarly abstracts in pharmacology, sociology, and linguistics and literature. *Journal of the American Society for Information Science and Technology* 2008;59(9):1420–1432.
- 11 Ufnalska S. Abstracts of research articles: problems of translation. *European Science Editing* 2007;33(4):101–104.
- 12 Młyniec W, Ufnalska S. *Scientific communication, czyli jak pisać i prezentować prace naukowe*. Poznań: Wydawnictwo Sorus, 2003.
- 13 Swales JM, Feak CB. *Abstracts and the writing of abstracts*. Michigan: University of Michigan Press, 2009.