

Readability in the *British Journal of Surgery*

J. D. Hayden

Department of Surgery, D Floor, Clarendon Wing, The General Infirmary at Leeds, Leeds LS1 3EX, UK
Correspondence to: Mr J. D. Hayden (e-mail: jezhayden@googlemail.com)

Background: Readability scores predict the ease with which a document can be read and understood. Identification of the factors that affect readability might improve the quality of surgical literature.

Methods: Electronic copies of submitted, peer reviewed and edited manuscripts of 189 articles published in the *British Journal of Surgery* from March 2006 to April 2007 were studied. Flesch reading ease scores were calculated for each version. Type of article (meta-analysis, review, randomized trial or other original paper), first language of principal author, number of peer revisions and editor were recorded.

Results: Flesch score varied according to type of article ($P = 0.004$). The mean readability score was lower when English was the first language of the principal author ($P = 0.016$). There was no significant difference in mean Flesch score between the submitted and accepted (peer reviewed) manuscripts, but a significant increase occurred after editing ($P < 0.001$), which did not vary between editors.

Conclusion: Several factors influence readability. Knowledge of such factors might help authors to improve their scientific writing.

Paper accepted 22 November 2007

Published online in Wiley InterScience (www.bjs.co.uk). DOI: 10.1002/bjs.5994

Introduction

Readability refers to the ease with which a piece of text is understood by a particular class of people¹. Factors that influence readability include typography, legibility, individual differences (in for example comprehension or motivation), layout and use of illustrations¹. Readability may be estimated using a predictor score such as the Flesch reading ease scale², which is calculated from simple text measurements (number of words, sentences and syllables)³. This score can be determined using word processing software and has been used to predict readability in medical and nursing journal articles^{4,5}, consent forms⁶ and web-based information resources⁷.

Journal type and author nationality are associated with differences in readability score⁸⁻¹⁰. Although it is recognized that the processes of peer review and editing can increase the quality of articles in scientific journals^{4,9,11,12}, little is known about the factors that influence the readability of surgical articles. Published instructions to authors tend to provide information about formatting rather than readability¹³. A recent study of

articles published in three surgical journals suggested that they contained too many large words and long sentences, and that their readability could be improved¹⁴. The present study assessed submitted, accepted (peer reviewed) and final version (edited and copy edited) manuscripts in order to determine the variables that affect the readability score of articles published in the *British Journal of Surgery* (*BJS*).

Methods

Manuscript sources

Manuscript Central is an internet-based tool that processes articles submitted to biomedical journals¹⁵. *BJS* started to use Manuscript Central in March 2006, and from then to the end of April 2007, 244 articles were published in the journal. Rejected manuscripts were not considered in this study.

Using Manuscript Central with Content Editor access privileges, digital copies of submitted and accepted (after revision) versions of 189 articles were retrieved editorials and leaders were excluded. The final (edited) manuscript for each article was obtained electronically from the publisher's website¹⁶. For each version, the words from the introduction, methods, results and discussion were cut

The Flesch score of this manuscript at submission was 23.4. This changed to 25.7 after the first revision and became 32.0 after editing.

and pasted into a separate blank document as unformatted text and saved. Errors that appeared during downloading of hypertext mark-up language were corrected before analysis.

Two assumptions were made in this study. First, it was assumed that the peer review process influenced changes in manuscript content between submission and acceptance. Second, it was assumed that the editing process (including copy editing) influenced changes in manuscript content between acceptance and publication.

Text measurements

The Flesch reading ease score was calculated using the Microsoft® Word spelling and grammar tool (Microsoft, Redmond, Washington, USA). The calculation is based on the equation $206.835 - 1.015 \times (\text{total words}/\text{total sentences}) - 84.6 \times (\text{total syllables}/\text{total words})^3$. The maximum score is 121 and there is no lower limit. The Flesch score was chosen for simplicity and convenience of measurement. The higher the score, the more readable the text is predicted to be. Other scoring systems are available that produce an index referring to the grade of student from a USA school that should be able to understand the text^{17,18}. Predicted readability has an inverse relationship with these scores.

Text measurements of total words, paragraphs, sentences and characters, as well as mean number of sentences per paragraph, words per sentence and characters per word and number of passive sentences were performed using the Microsoft® Word spelling and grammar tool. Other recorded variables included the type of article, country of origin, the first language of the principal author, the number of revisions before acceptance and which of six editors had been assigned to the manuscript.

Statistical analysis

Text calculations and variable scores were entered into a Microsoft® Excel spreadsheet (Microsoft) and imported into SPSS® version 11 (SPSS, Chicago, Illinois, USA) for analysis. Means of continuous variables were compared using the paired- or independent- samples *t* test and one-way analysis of variance (ANOVA) as appropriate. $P < 0.050$ was considered significant. Data are presented as means with 95 per cent confidence intervals.

Results

Submitted manuscripts

Of the 189 articles, nine (4.8 per cent) were meta-analyses, 27 (14.3 per cent) were reviews, 34 (18.0 per cent) were

randomized clinical trials and 119 (63.0 per cent) were other types of original article. The geographical origin of the manuscripts was Europe for 145 (76.7 per cent), the USA and Canada for ten (5.3 per cent), South East Asia for 21 (11.1 per cent) and Australasia for 13 (6.9 per cent). English was the first language of the principal author of 77 manuscripts (40.7 per cent).

There was no significant difference in the mean Flesch score relating to geographic origin ($P = 0.168$). However, manuscripts written by an author whose first language was English had a significantly lower mean Flesch score than those written by an author for whom English was a foreign language ($P = 0.016$) (Fig. 1).

The mean Flesch score also varied significantly according to manuscript type ($P = 0.004$). Original articles and randomized clinical trials generally scored better than meta-analyses and reviews (Fig. 2).

Peer review

Submitted manuscripts had been assigned to one of six editors and sent for peer review. Nine (4.8 per cent) of the 189 manuscripts were accepted without revision; 133 (70.4 per cent) required a single revision and 47 (24.9 per cent) required two or more revisions before acceptance. There was no significant difference in mean Flesch score between submitted and accepted (peer reviewed) manuscripts ($P = 0.435$) (Fig. 3). Change in score at this stage was assessed according to the number of

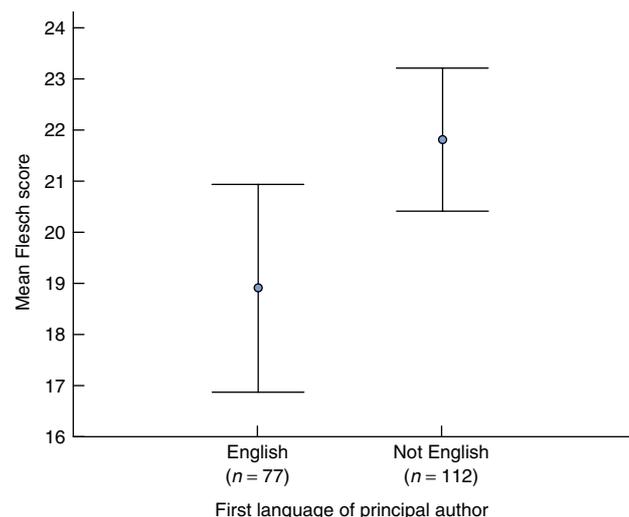


Fig. 1 Mean Flesch scores for submitted manuscripts according to the first language of the principal author, with error bars representing 95 per cent confidence intervals. $P = 0.016$ (independent- samples *t* test)

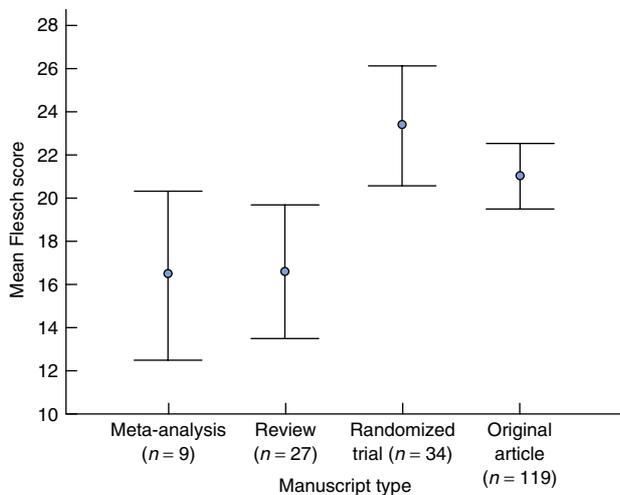


Fig. 2 Mean Flesch scores for submitted manuscripts according to type of article, with error bars representing 95 per cent confidence intervals. $P = 0.004$ (one-way ANOVA)

revisions. Manuscripts that received two or more revisions had a significantly lower mean change in readability score ($P = 0.028$). The only significant difference in text variable between the manuscripts at the peer review stage of publishing was in the mean proportion of passive sentences ($P < 0.001$) (Table 1).

Editing

The mean Flesch score significantly improved between manuscript acceptance and publication ($P < 0.001$) (Fig. 3). The mean increase was 16.1 per cent (Table 1). The mean numbers of words, characters, paragraphs, sentences, sentences per paragraph and characters per word significantly decreased between the accepted and published manuscripts (Table 1). When this mean change in score was analysed by editor, no significant difference was found ($P = 0.649$). Examples of text from published original articles with the highest and lowest recorded Flesch scores appear in Fig. 4.

Discussion

In this study, Flesch scores indicated that readability was lower when the principal author's first language was English. In addition, meta-analyses and reviews were generally less readable than randomized clinical trials and other types of original article. Although peer review made no difference to the readability score, the editing stage significantly improved it, with no significant difference between editors.

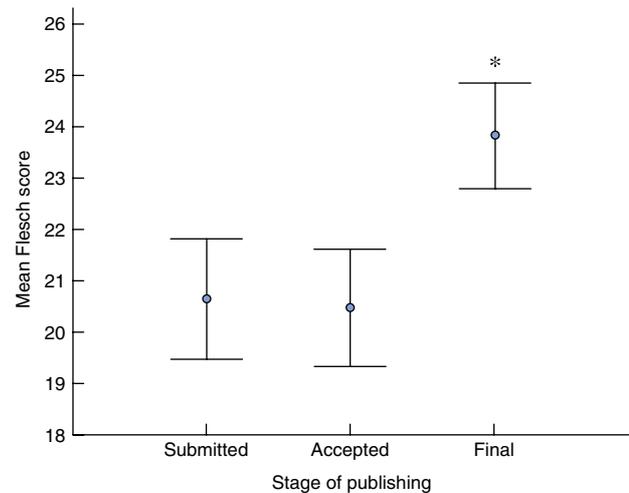


Fig. 3 Mean Flesch scores of 189 manuscripts at different stages of publishing, with error bars representing 95 per cent confidence intervals. * $P < 0.001$ versus accepted manuscript (paired-samples t test)

Editors aim to increase the impact factor of their journals¹⁹, but improving readability has received less attention^{4,5}. Authors receive hardly any instruction on this¹³, and papers are rarely rejected by editors because of problems with readability^{20,21}.

This study showed that manuscripts submitted by authors whose first language was English were less readable than those from authors for whom English was a foreign tongue. The latter probably constructed shorter, less complex, sentences. Some may have used an independent professional manuscript-writing service. These factors may affect the readability of surgical journals. Furthermore, most articles published in surgical journals come from the USA²², and readability scores of articles by American authors tend to be lower than those of British writers²³. In this context, it is interesting that the proportion of articles published by European and non-American authors is increasing. Perhaps readability will improve²⁴.

Lower readability scores for reviews and meta-analyses is a new observation; the literature lacks evidence to confirm this. Systematic reviews and meta-analyses occupy the highest echelons of research evidence, presenting a rigorous evaluation of the published literature²⁵. However, these articles are often unpopular with readers²⁶. Whether their poor readability translates into poor understanding by readers is not known.

Peer review is a ubiquitous yet controversial part of scientific publishing. The process is slow, expensive, subject to bias and has little evidence base^{27,28}. Nevertheless, a combination of peer review plus editing is known

Table 1 Summary of text values for manuscripts at each stage of publishing

	Submitted		Accepted		Final
	Mean	P*	Mean	P†	Mean
Words	2723 (2599, 2847)	0.719	2736 (2608, 2864)	< 0.001	2524 (2403, 2644)
Characters	15 262 (14 553, 15 970)	0.712	15 336 (14 604, 16 067)	< 0.001	13 923 (13 245, 14 601)
Paragraphs	38.6 (36.6, 40.5)	0.985	38.6 (36.6, 40.5)	< 0.001	33.9 (32.4, 35.4)
Sentences	114.4 (109.2, 119.6)	0.834	114.1 (108.8, 119.3)	< 0.001	108.4 (103.4, 113.4)
Sentences per paragraph	4.6 (4.5, 4.8)	0.657	4.6 (4.4, 4.8)	0.045	4.5 (4.4, 4.6)
Words per sentence	23.3 (22.8, 23.7)	0.100	23.4 (22.9, 23.9)	0.138	23.1 (22.8, 23.5)
Characters per word	5.4 (5.4, 5.5)	1.000	5.4 (5.4, 5.5)	< 0.001	5.3 (5.3, 5.3)
Passive sentences	37.0 (35.8, 38.2)	< 0.001	38.1 (36.9, 39.2)	0.103	38.7 (37.6, 39.8)
Flesch score	20.6 (19.5, 21.8)	0.435	20.5 (19.3, 21.6)	< 0.001	23.8 (22.8, 24.9)

Values in parentheses are 95 per cent confidence intervals. **Versus* accepted manuscript; †*versus* final manuscript; paired-samples *t* test.

Lowest ranking score	Highest ranking score
<p>In this study anastomotic leak occurred in 6.4 per cent of patients who underwent hepatobiliary resection involving an intrahepatic cholangiojejunostomy. Several authors have reported an incidence of anastomotic leak of 0.4–8 per cent after hepaticojejunostomy distal to the hepatic hilum in pancreatic resection or bypass surgery. De Castro <i>et al.</i> recently reviewed details of their experience with hepaticojejunostomy in 1033 patients (distal to the hepatic hilum in 989, proximal to the hepatic hilum in 44), following which anastomotic leak of the hepaticojejunostomy occurred in 2.3 per cent. Only the study of Miyazaki <i>et al.</i> appeared to include a high proportion of intrahepatic cholangiojejunostomies; anastomotic breakdown occurred in 22 per cent (14 of 65) of patients who underwent hepatectomy for hilar cholangiocarcinoma, but no specific details about the difficulty of anastomosis were given. Other authors mentioned bile leakage in 1.0–21.8 per cent of patients with hilar cholangiocarcinoma who underwent resection with or without hepatectomy. These reports, however, did not clearly specify whether bile leakage arose from the liver stump or from anastomotic leak of the hepaticojejunostomy. In the absence of detailed reports explicitly focusing on intrahepatic cholangiojejunostomy, there is no standard against which the present results can be evaluated. Considering the technical difficulties and highly invasive surgery needed, an overall incidence of anastomotic leak of 6.4 per cent appears acceptable.</p> <p>Final manuscript score, 8.6; paragraph score, 0</p>	<p>A consecutive series of patients whose chronic leg ulcers failed to heal with prolonged conservative management were treated by wide excision and meshed split-skin grafting. All but three were healed on discharge and more than half of these ulcers remained healed for up to 5 years. Ulcer recurrence was most likely to occur in the first 2 months after grafting, in keeping with other series. Most ulcers that were healed at 2 months in the study remained healed at 5 years. There was no control group, in whom conservative management was continued, but all patients in the series had been treated previously with at least 6 months of compression bandaging. The present study also demonstrated that ulcers of various aetiologies can be healed successfully and that the breakdown rate was similar for all types of ulcer. The size of the ulcer did not affect the healing rate, again confirming the findings of others.</p> <p>Final manuscript score, 40.4; paragraph score, 55.2</p>

Fig. 4 Example discussion paragraphs from articles with the lowest and highest ranking Flesch scores

to improve the quality of published articles^{4,5,9,11}. A meta-analysis has also demonstrated a slight improvement in readability through 'technical editing' that included peer review²⁹. Recently, a systematic Cochrane review confirmed that the overall package of technical editing improves paper quality³⁰. However, the present study found no significant difference in readability score attributable to peer review; indeed, a negative effect on readability was noted if there were two or more revisions after peer review. The significance of this is uncertain, but it raises the suspicion that repeated adjustment of text adversely affects readability.

Attempts have been made to improve the peer review process. Structured training has proved unpopular³¹ and,

although a randomized trial assessing a self-improvement course showed some increase in review quality, this was not of editorial significance³². Others have shown that a panel of dedicated methodology and statistical reviewers failed to improve manuscript quality³³. In this study, editors and authors ignored half of all suggestions for manuscript improvement³³. Neither masking the authors nor revealing the identity of reviewers has made any difference to the quality of peer review^{32,34}.

Editing was identified as making a significant impact on readability. This result is novel as others have studied differences in readability scores only between submitted and final manuscripts without separating 'technical editing' into components^{4,5}. Reassuringly, there was no significant

difference between editors with respect to improvement in score. The editing process is complex, involving interaction between editors, referees and authors. Quality control of the final manuscript includes adjustments made by copy editors. Editing has been shown to improve the quality of abstracts and accuracy of reference lists²⁹. It also involves dealing with too much, too little, inaccurate and misplaced information, as well as problems with manuscript layout³⁵. The present study identified significant reductions in most of the recorded text measurements after the editing stage of publishing. Most of these changes reduce manuscript length. The editing process reduced the mean word count of manuscripts by 7.8 per cent. It would be interesting to know if asking authors to shorten manuscript length by a similar or greater amount would achieve similar changes in readability score.

Some have criticized the Flesch score as a predictor of readability³⁶. None of the available score systems has been specifically designed for use with scientific literature. However, the Flesch score is being used increasingly in this context and has been validated^{4,5,14,23}. Still, the present study used only a single readability measurement. It is possible that other scoring systems, such as the Coleman–Liau index or the Simple Measure of Gobbledygook score, might produce different results^{17,18}.

In the light of the above, it would seem reasonable to provide authors with information on readability and how to improve it. This might benefit surgical publishing. Hall¹⁴ has suggested that authors aim for a Flesch score of over 30 to avoid the possibility of their manuscripts being as difficult to read as legal documents. While this level of text might be beyond the surgical authorship, or even inappropriate for surgical work, perhaps it is something to aspire to.

Acknowledgements

The author thanks Karen Shield and Bryony Urquhart of John Wiley and Sons for facilitating access to manuscript archives, and John Murie and Jonathan Earnshaw, editors of *BJS*, for their support.

References

- McLaughlin GH. Proposals for British readability measures. In *The Third International Reading Symposium*, Downing J, Brown AL (eds). Cassell: London, 1968; 186–205.
- Flesch R. A new readability yardstick. *J Appl Psychol* 1948; **32**: 221–233.
- Flesch–Kincaid Readability Test. http://en.wikipedia.org/wiki/Flesch–Kincaid_Readability_Test [accessed 22 March 2007].
- Biddle C, Aker J. How does the peer review process influence AANA journal article readability? *AANA J* 1996; **64**: 65–68.
- Roberts JC, Fletcher RH, Fletcher SW. Effects of peer review and editing on the readability of articles published in *Annals of Internal Medicine*. *JAMA* 1994; **272**: 119–121.
- Beyer DR, Lauer MS, Davis S. Readability of informed-consent forms. *N Engl J Med* 2003; **348**: 2262–2263.
- Friedman DB, Hoffman-Goetz L, Arocha JF. Readability of cancer information on the internet. *J Cancer Educ* 2004; **19**: 117–122.
- Poulton EC. Effects of printing types and formats on the comprehension of scientific journals. *Nature* 1959; **184**(Suppl 23): 1824–1825.
- Goodman SN, Berlin J, Fletcher SW, Fletcher RH. Manuscript quality before and after peer review and editing at *Annals of Internal Medicine*. *Ann Intern Med* 1994; **121**: 11–21.
- Hedl JJ Jr, Glazer-Waldman HR, Parker HJ, Hopkins KM. Readability and writing style analysis of selected allied health professional journals. *J Allied Health* 1991; **20**: 25–37.
- Pierie JP, Walvoort HC, Overbeke AJ. Readers' evaluation of effect of peer review and editing on quality of articles in the *Nederlands Tijdschrift voor Geneeskunde*. *Lancet* 1996; **348**: 1480–1483.
- Hochhauser M. Readability of British and American medical prose. Why are unreadable articles still being written? *BMJ* 2003; **326**: 711.
- Schriger DL, Arora S, Altman DG. The content of medical journal Instructions for authors. *Ann Emerg Med* 2006; **48**: 743–749, 749e.741–744.
- Hall JC. The readability of original articles in surgical journals. *ANZ J Surg* 2006; **76**: 68–70.
- Manuscript Central. <http://mc.manuscriptcentral.com/bjs> [accessed 22 March 2007].
- Wiley Interscience. <http://www3.interscience.wiley.com/cgi-bin/home> [accessed 22 March 2007].
- Coleman M, Liau TL. A computer readability formula designed for machine scoring. *J Appl Psychol* 1975; **60**: 283–284.
- McLaughlin GH. SMOG grading: a new readability formula. *J Reading* 1969; **12**: 639–646.
- Chew M, Villanueva EV, Van Der Weyden MB. Life and times of the impact factor: retrospective analysis of trends for seven medical journals (1994–2005) and their Editors' views. *J R Soc Med* 2007; **100**: 142–150.
- Pierson DJ. The top 10 reasons why manuscripts are not accepted for publication. *Respir Care* 2004; **49**: 1246–1252.
- Radford DR, Smillie L, Wilson RF, Grace AM. The criteria used by editors of scientific dental journals in the assessment of manuscripts submitted for publication. *Br Dent J* 1999; **187**: 376–379.
- van Rossum M, Bosker BH, Pierik EG, Verheyen CC. Geographic origin of publications in surgical journals. *Br J Surg* 2007; **94**: 244–247.

- 23 Weeks WB, Wallace AE. Readability of British and American medical prose at the start of the 21st century. *BMJ* 2002; **325**: 1451–1452.
- 24 Tompkins RK, Ko CY, Donovan AJ. Internationalization of general surgical journals: origin and content of articles published in North America and Great Britain from 1983 to 1998. *Arch Surg* 2001; **136**: 1345–1351.
- 25 Akobeng AK. Understanding systematic reviews and meta-analysis. *Arch Dis Child* 2005; **90**: 845–848.
- 26 Loke YK, Derry S. Does anybody read 'evidence-based' articles? *BMC Med Res Methodol* 2003; **3**: 14.
- 27 Goldbeck-Wood S. Evidence on peer review – scientific quality control or smokescreen? *BMJ* 1999; **318**: 44–45.
- 28 Callaham ML. Research into peer review and scientific publication: journals look in the mirror. *Ann Emerg Med* 2002; **40**: 313–316.
- 29 Wager E, Middleton P. Effects of technical editing in biomedical journals: a systematic review. *JAMA* 2002; **287**: 2821–2824.
- 30 Wager E, Middleton P. Technical editing of research reports in biomedical journals. *Cochrane Database Syst Rev* 2007; (2)MR000002.
- 31 Callaham ML, Schriger DL. Effect of structured workshop training on subsequent performance of journal peer reviewers. *Ann Emerg Med* 2002; **40**: 323–328.
- 32 Schroter S, Black N, Evans S, Carpenter J, Godlee F, Smith R. Effects of training on quality of peer review: randomised controlled trial. *BMJ* 2004; **328**: 673.
- 33 Schriger DL, Cooper RJ, Wears RL, Waeckerle JF. The effect of dedicated methodology and statistical review on published manuscript quality. *Ann Emerg Med* 2002; **40**: 334–337.
- 34 van Rooyen S, Godlee F, Evans S, Black N, Smith R. Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial. *BMJ* 1999; **318**: 23–27.
- 35 Purcell GP, Donovan SL, Davidoff F. Changes to manuscripts during the editorial process: characterizing the evolution of a clinical paper. *JAMA* 1998; **280**: 227–228.
- 36 Ludbrook J. Readability of original articles in surgical journals. *ANZ J Surg* 2006; **76**: 1038–1040.

If you wish to comment on this, or any other article published in the *BJS*, please visit the on-line correspondence section of the website (www.bjs.co.uk). Electronic communications will be reviewed by the Correspondence Editor and a selection will appear in the correspondence section of the Journal. Time taken to produce a thoughtful and well written letter will improve the chances of publication in the Journal.
