Improving access to COVID-19 information by ensuring the readability of government websites

Tanya Serry | Tonya Stebbins | Andrew Martchenko | Natalie Araujo | Brigid McCarthy

1School of Education, La Trobe University, Bundoora, Victoria, Australia
2School of Humanities and Social Sciences, La Trobe University, Melbourne, Victoria, Australia
3La Trobe University, Melbourne, Victoria, Australia

Correspondence
Tanya Serry, School of Education, La Trobe University, Melbourne Campus, Bundoora VIC 3086, Australia.
Email: t.serry@latrobe.edu.au

Abstract

Issue addressed: This study evaluated the readability of web pages from two public-facing Victorian government websites that were responsible for communicating key health messages relating to the COVID-19 pandemic in 2020.

Methods: Webpages were downloaded and filtered to identify relevant materials (English language materials containing HTML files that referred to COVID-19). The files were converted to text files and two Python packages, SpaCy and TextStat were used to obtain the data presented here. In addition to running two well-established readability tests, SMOG Index (Simple Measure of Gobbledygook) and Flesch Reading Ease formula, we also calculated the figures for sentence length and word length, which drive the readability measures and allow a disaggregated view of the data. Type token ratio measures were conducted as a reflection of the breadth of vocabulary used in the web pages.

Results: Derived measures of text complexity were higher than recommended levels of text complexity for health promotion materials, which are generally set at senior primary school levels. This did not vary depending on the intended audience (public or professional). A senior secondary reading level was required for effective engagement with the text published on both sites.

Conclusions: Improving the readability of materials on key government websites where information about COVID-19 is being communicated to the public, represents a low cost and potentially effective means of improving public understanding of the pandemic and the steps individuals need to take to protect themselves and the community.

So what?: Given the challenges widely identified in ensuring compliance with protective behaviours, confidence in seeking vaccination and increasing distrust of government, it would be strategic to improve public communication to ensure health messages are simple and readily understood.

Summary: The complexity and readability of text contained in web pages during 2020 from two Victorian government departments were evaluated. Communication regarding the restrictions and the management of risks associated with COVID-19
was the main focus of these 367 individual web pages from the Department of Health and Human Services (DHHS) and the Department of Education and Training (DET). Results indicated that across both sites and on both readability measures used, an education level equivalent to senior secondary school would be required to readily understand the contents.

KEYWORDS
COVID-19, Flesch, plain English, readability, SMOG, website

1 | INTRODUCTION

Effective communication with the public is a critical component for managing the health and well-being of the community, particularly during the current COVID-19 pandemic. There have been persistent challenges with clear and accessible communication during this pandemic on several fronts, including the constant updating of advice, the proliferation of misinformation and lack of accessible content for a range of vulnerable populations such as those experiencing social disadvantage (for example, in Victoria, 13.2% live in poverty and 26% of Victorians identify a language other than English as their household language). Further, it has been suggested that a lack of readability of key COVID-19-related public health messaging may even contribute to vaccine hesitancy. As Smith and Judd point out, governments struggle to enact policies that address the social determinants of low health literacy and access to care, and this has been frequently documented, particularly in the health sector. These long-standing issues are becoming increasingly visible in new ways because of the pandemic. During health crises, structural barriers and negative emotional states such as anxiety tend to further reduce the capacity of readers to process information effectively. For this reason, it is timely to consider both how these inequities can be addressed over the medium and long term but also, what immediate steps can be taken to increase access to critical information across the population.

The need for accessible public-facing information is reinforced by the Plain English Foundation, which reports that on average, Australian public sector documents are written at readability levels for those who have attained a post-secondary level of education. The Australian Federal Government has endorsed accessibility guidelines that set out standards for written communication to the public, stating that communications should be written to suit a lower secondary level of education. This equates to students at around Year 7. The Federal Government’s advice aligns with research from the Australian Bureau of Statistics in 2013, showing that over 40% of the population have literacy levels below those who have completed secondary school. As such, an apparent mismatch exists between recommended guidelines for accessible written communications for the public and what has typically been produced by government departments.

We argue that one of the simpler actions that governments could take is to further consider the text complexity of written communications to ensure that materials are easily accessible to the public. This can be achieved easily and efficiently by applying a formal measure of readability to written texts. Readability measures, of which there are many, have been used for decades to indicate the complexity of written language for newspapers and other publicly available materials. Regardless of the specific readability measure selected, the result is typically expressed as a number reflecting the years of formal education required to read and understand written material. Readability measures are based on word and sentence length contained within a text to provide an “objective estimate” about the demands of the text in relation to the reader’s educational level. Whilst word and sentence length metrics are only one aspect of clear and transparent written text, evidence suggests that compliance to objective measures of readability is one way of ensuring that critical information is accessible for much of the population.

Our research evaluates the readability of web-based materials about COVID-19 published on two key Victorian government websites in 2020. The state of Victoria provides an important case study. The Victorian government was the most prolific state communicator of COVID-19 information during the early stages of the pandemic and endured the most restrictive and sustained set of restrictions in Australia during this period. The quantity and breadth of information communicated offer insights into the challenges of providing timely and accessible information in rapidly changing contexts. To investigate the readability of this information, the websites selected were the Department of Health and Human Services (DHHS) and the Department of Education and Training (DET) as these are key sites for public-facing communications about health and education services. Our specific research aims were as follows:

1. To describe the readability measure results for public-facing DHHS and DET web pages during 2020.
2. To determine whether readability measures differed for web pages that appeared to address a general audience compared to those directed to workers in the health (DHHS) or education (DET) sectors with a range of levels of training and qualifications.

This paper contributes to an emerging body of evidence that examines the readability of web-based COVID-19 information. This literature is diverse but has tended to focus on cross-country comparisons of official information distributed by national and international health bodies, commonly sought information available on the
Internet, information related to specific behaviours such as vaccination, physical distancing, or mask wearing, or translated and easy English materials targeted at specific audiences. These approaches offer comparative breadth to understanding readability issues during the COVID-19 pandemic. Our focus on communications from the DHHS and DET during 2020 seeks to contribute to these conversations by providing an in-depth examination of readability within a temporally and geo-spatially defined information ecology. This study considers the full range of COVID-19 communications produced by DHHS and DET for Victorian consumers during 2020 and offers insights into the population-level accessibility of these communications.

2 | METHODS

2.1 | Method of data collection

The process of acquiring and preparing data from the DHHS and DET websites for analysis involved the following series of steps:

1. Download websites.
2. Select HTML files.
3. Select English pages.
4. Select content identified as the raw HTML text.
5. Select pages containing key content.
6. Convert URLs to readable text to facilitate word counts.
7. Convert to text files and test.

The tasks and tools associated with each of these steps are described as follows. The Linux tool wget was used to download all the contents on the DHHS and DET websites. The downloaded data contained HTML files as well as other types of documents/images. In this step, we recursively searched all website directories and select only HTML files. We then selected English pages by looking at the “en” attribute within the HTML elements and selecting pages where the value is “en” (English). The next step was to focus on the content in the pages and exclude invisible or irrelevant material such as the content of sidebar menus and so on. This was achieved by inspecting the HTML text and identifying the set of tags that contain content. The python package Beautiful Soup was used to parse the HTML files to select the text in these tags. The next task was to exclude pages that were not focussed on COVID-19. To do this, we searched for pages containing three or more occurrences of the keywords “COVID” or “coronavirus” and excluded the remaining pages from the analysis. Generally, a single mention of the keywords was associated with a sentence alerting users to other pages containing information about COVID-19. Word counts formed an important part of the analysis, so it was necessary to convert URLs to readable text. This was accomplished by using Regular Expressions to find text containing websites and email text patterns (www, http, name@company.com...) and replacing these sequences with readable text (“HYPERLINK” or “EMAIL”). Finally, the data were converted to text files using Beautiful Soup. One additional change that is worth noting here is that full stops were inserted into headings and lists to counteract an inflation of the length of sentences when full stops were omitted. Given the considerable length of many list items in the corpus and the significance of headings for readability, this was judged to be a better reflection of reader engagement in the documents than omitting this material or treating it as single sentences, which would have inflated the reading scores where sentence length is part of the measure.

2.2 | Method of data analysis

A range of different variables relating to the complexity of the language used within the sites were analysed to assess the readability of the text documents from the data collection process. Some of these are built into readability tests and some were identified as providing additional useful information. Examples included: (i) the number of unique words on a web page, (ii) a sentence count and (iii) a tally of the number of clauses on a web page. The variables reported here are generated through two Python packages: SpaCy is used to count unique words and lemmas, whilst TextStat produces the remaining results.

3 | RESULTS

A total of 425 web pages were gathered across the DHHS website (n = 386) and the DET website (n = 39). Once duplicates were removed, our data set comprised 348 individual pages from the DHHS and 19 pages from DET. Pages from each site were then filtered by the first two researchers to extract those that appeared to be specifically aimed at health or education professionals respectively, along with the majority of pages that were deemed to be public facing. The determination was based on the content of the pages and was often apparent in their titles. On the DHHS website, we determined that nine of the 348 pages were targeting health professionals. Examples include pages titled (i) GP Respiratory Clinics and Hospital Respiratory Clinics (COVID-19) and (ii) For service providers - coronavirus (COVID-19). For the DET website, we identified nine of the 19 pages targeting educational professionals from early childhood to college level. Examples were pages titled (i) Community Child Care Fund Special Circumstances Grant Opportunity and (ii) COVID-19 - Resources for teachers and school leaders. Figure 1 displays the data extraction and filtering process.

From the multiple results generated using the tools SpaCy and TextStat, we derived six measures that reflect the complexity and density of the text on the web pages. These are shown in Table 1. They include two well-established readability tests, the Simple Measure of Gobbledygook (SMOG Index) and the Flesch Reading Ease (FRE) formula. In addition, we include measures of sentence length and word length. Word and sentence length drive the readability tests and including them allow for a disaggregated view of the data. We also calculate the Type-Token Ratios as an indicator of the range of
Similarly, the median and Web Content indicates that an average Derived measures used for data analysis recommendations is around 7. (As the number increases, readability becomes more challenging). As shown in Table 2, the derived measures are generally similar across both sites. For example, the median number of words per sentence for DHHS pages was 13.44, whilst the median for the DET pages was 14.44. For the two formal readability measures, the median FRE Formulae were 60.65 (DHHS) and 52.26 (DET). These equate to an educational level of close to Grade 10 or higher which refers to “fairly difficult” text (see p. 26). Similarly, the median SMOG indices were 11.45 (DHHS) and 12.20 (DET) which reflects reading comprehension at senior secondary levels (see p. 644).

Figures 2 and 3 display box and whisker plots for the median measure and the interquartile ranges for the FRE Formula (Figure 2) and the SMOG index (Figure 3). These data are presented in relation to the Australian Federal Government’s endorsed accessibility guidelines for public-facing information.

Shaded grey areas refer to the FRE scores in the recommended readability range for written communication in the public domain. FRE scores falling in the dark-shaded area reflect the educational level of Year 6 and pale shared area reflects the educational level of Year 7.

Scoring Key: 0-30 = college graduate; 30-50 = college student; 50-60 = students in Years 10-12; 60-70 = students in Years 8-9; 70-80 = students in Year 7; 80-90 = students in Year 6 (Recommended readability level) and 90-100 = students in Year 5.

Shaded grey areas refer to the SMOG Index scores in the recommended readability range for written communication in the public domain. A SMOG index falling in the dark-shaded area reflects the educational level of Years 5-6 and pale shared area reflects the educational level of Years 7-8.

Scoring Key: 10+ = easily understood by average college graduate; 9-9.9 = easily understood by average undergraduate tertiary students; 8-8.9 = easily understood by average Years 11-12 students; 7-7.9 = easily understood by average Years 9-10 students; 6-6.9 = easily understood by average Years 7-8 students; 5-5.9 = easily understood by average Years 5-6 students.

Flesch 1949 identifies 1.39 syllables per 100 words as “fairly easy” reading because it signals that a rather constrained vocabulary is being used. Measures from our websites reflected a higher syllable to word ratio ranging from 1.55 to 1.64 which is characterised as “fairly difficult.” Interestingly, Flesch 1949 indicates that an average of 14 words per sentence is in the “fairly easy”, upper primary range.

This is consistent with our results, as indicated in Table 2. However, due to the need to manage dot point lists in our data set, by inserting

---

TABLE 1  Derived measures used for data analysis

| Derived measure                  | Calculation process                                                                 |
|----------------------------------|-------------------------------------------------------------------------------------|
| Syllables per word               | The total number of words as a ratio of the total number of syllables                |
| Words per sentence               | The total number of sentences as a ratio of the total number of words                |
| Type-Token Ratio                 | The total number of unique words (inclusive of all stop words) as a ratio of the total number of words |
| Clausal density                  | The total number of clauses as a ratio of the total number of sentences               |
| SMOG Index (Simple Measure of Gobbledygook) | A readability measure designed to estimate the education level required for an individual to read and comprehend text with particular attention to word length. The score consistent with a Year 7 reading level as indicated by the Australian government and Web Content Accessibility Guidelines (WCAG) recommendations is around 7. (As the number increases, readability becomes more challenging) |
| Flesch Reading Ease (FRE) formula | A readability measure also designed to estimate the education level required for an individual to read and comprehend text, with attention to both word length and sentence length. The score consistent with a Year 7 reading level as indicated by the Australian government and WCAG recommendations is around 75. As the number decreases, readability becomes more challenging |

vocabulary required to understand the text. Finally, we present a measure of clausal density, which provides a measure of the complexity of the sentences within the text.

Data are presented according to the DHHS and the DET sites separately. As shown in Table 2, the derived measures are generally similar across both sites.
sentence breaks between items, there was a reduction in the overall measure of words per sentence in our data. For this reason, we relied on the integrated readability measures rather than considering the disaggregated measures in more detail.

To determine whether the linguistic complexity differed for web pages that appeared to be professional facing or public facing, we calculated medians of our derived measures for both. We chose to report the median rather than the mean to account for one very large web page from the DET.

Table 3 illustrates that there was little difference in the complexity and density of the measures when the pages were separated by the projected audience.

**Table 2.** Derived measures of text complexity and density of text for the DHHS and the DET websites

|                      | Syllables per word | Words per sentence | Type-token ratio | Clausal density | SMOG  | FRE formula |
|----------------------|--------------------|--------------------|------------------|----------------|-------|-------------|
| **Department of Health and Human Services (n = 348)** |                    |                    |                  |                |       |             |
| Mean                 | 1.58               | 13.78              | 0.44             | 1.92           | 11.49 | 59.36       |
| SD                   | 0.11               | 3.61               | 0.11             | 0.42           | 1.43  | 9.29        |
| Median               | 1.55               | 13.40              | 0.44             | 1.95           | 11.45 | 60.65       |
| Minimum              | 1.36               | 2.82               | 0.04             | 1.00           | 8.30  | 20.89       |
| Maximum              | 2.03               | 26.68              | 0.84             | 4.00           | 16.40 | 81.49       |
| **Department of Education and Training (Victoria) (n = 19)** |                    |                    |                  |                |       |             |
| Mean                 | 1.64               | 14.71              | 0.44             | 2.32           | 12.49 | 52.99       |
| SD                   | 0.11               | 4.30               | 0.14             | 0.73           | 1.61  | 11.52       |
| Median               | 1.66               | 14.44              | 0.41             | 2.17           | 12.20 | 52.26       |
| Minimum              | 1.44               | 9.05               | 0.09             | 1.57           | 10.30 | 34.52       |
| Maximum              | 1.88               | 23.08              | 0.66             | 4.00           | 15.60 | 73.07       |

Abbreviations: FRE, Flesch Reading Ease formula; SMOG, Simple Measure of Gobbledygook.

**Figure 2.** Flesch Reading ease median scores for DHHS and DET websites (n = 386)

**Figure 3.** SMOG index median scores for DHHS and DET websites (n = 386)

Our results show that the web page contents from the DHHS and DET in 2020 rarely met federally endorsed accessibility guidelines. As a result, these contents are likely to be inaccessible to a sizeable portion of the population. Our findings replicate a recent Polish study of 61 COVID-19-related health literacy web page contents in which none of the articles were suitably targeted for a 5th or 6th grade educational level.

The level of complexity in the text has both immediate and cumulative effects. Complex texts create an overall reading experience that
is likely to increase anxiety and exacerbate maladaptive behaviours.\(^1\) Better readability of these materials would increase understanding in the community, with potential flow on benefits in terms of compliance with restrictions, engagement with vaccination programs and everyday protective behaviours in a pandemic. Simpler source documents would also facilitate the translation of materials into other languages.

We acknowledge that readability measures are not foolproof because they reflect profile measures rather than descriptive measures of text. However, there is a consistent picture in our data that information on web pages represents complex text that would be inaccessible to many. The SMOG index\(^{25}\) and Flesch Reading Ease Formulae\(^{28}\) evaluate various aspects of text but generated equivalent results in our data, indicating that the readability reflects that a minimum senior secondary level of education would be required for the effective reading of key COVID-related government health and education communications.

Experience of the pandemic has demonstrated both the challenges of ensuring population-level compliance with protective measures such as testing, wearing masks and staying at home, and the ways in which risk is compounded in disadvantaged communities. Changing health advice exacerbates fear and confusion in the community. This is particularly the case when the information communicated challenges the accessibility thresholds of community members. In such circumstances, information may negatively impact trust both in officials and in the validity of the information presented, which may, in turn, negatively impact health-preserving and seeking behaviours (See Berg et al.\(^{31}\)). Effective communication is a key to reducing risk and supporting compliance.\(^7\)

Effort and skill are needed to write complex ideas in a straightforward way, and this challenge has been intensified in the pandemic by the need to communicate new concepts as a matter of urgency. Attending to readability is strategic in a context where significant amounts of information need to be communicated rapidly in a clear and transparent way. Readability measures are a cost-effective and easily implemented means to ensure that text about important topics is accessible.

### 5 | LIMITATIONS

There are some limitations in our research that must be acknowledged. Readability measures capture metrics involving word and sentence length but cannot shed light on more nuanced aspects such as grammaticality, text cohesion or clarity. Therefore, whilst readability measures are objective and have been used for many decades, they cannot reflect certain aspects of written text. To add the richness of our data, we derived measures such as clausal density and the ratio of novel words compared to the total number of words. Nevertheless, we acknowledge that qualitative descriptors of text could have added further depth to our analysis.

The layout of web pages was not considered in our analysis as our data were derived from text data only. The layout may or may not have enhanced the accessibility of web pages, but we did not evaluate this aspect. The software that we used to measure readability is not designed for evaluating web pages and therefore, certain features such as lists that appeared on various web pages required additional management. For example, in the research presented here, we managed extensive lists on some pages by consistently adding a full stop to the end of each item in any list.

Finally, our data were sourced from only two government department websites. However, we selected these two as the areas of health and education that generated vast quantities of critical COVID-related information which was published with the intention of managing risk across the population (DHHS) and in the education sector specifically (DET). In the Victorian context, these departments constituted the primary avenues for COVID-19 communication dissemination to the public in 2020.

### 6 | CONCLUSION

We analysed the readability of web pages sourced in 2020 from two Victorian government departments that were tasked with communicating and managing key information related to COVID-19. Two well-established measures of readability were used. Results showed that on average, readability scores indicated that a senior secondary level of education would be required for an effective understanding of the text. Careful monitoring of public messaging, using these tools, would be a simple and cost-effective measure for improving communications with the public. This is especially important in times of crisis such as the current COVID-19 pandemic. To facilitate best practice models,
we encourage the adoption of initiatives to improve the readability of government communications. In the first instance, we suggest that the use of free, publicly available websites to test the readability of public-facing communications circulated by DHHS and DET could improve readability and reduce risks. Whilst we recognise a time and labour cost associated with readability testing, we suggest that the incorporation of readability testing could ultimately reduce health risks such as nonadherence, increase trust in government communications and reduce long-term costs associated with crisis communication and management.

Our study has provided an in-depth analysis of the readability of Victorian DHHS and DET communications during the first year of the COVID-19 pandemic. It considers challenges to accessibility. In addition to procedural changes as mentioned above, we believe that this study shows the need for greater monitoring of the cost and care burdens associated with the circulation of information that exceeds population readability standards. Research into the costs of information adaption by intermediary organisations, is required. We argue that there is an urgent need for further research into the impacts of readability on the circulation and adaptation of COVID-19 health information at organisational and community levels. Taken together with our findings into the broad scope of readability challenges, this information could inform policy innovations.

ACKNOWLEDGEMENT

Thanks also to Alan Healey who worked on a previous project and thereby contributed to the methods and approach taken here. Open access publishing facilitated by La Trobe University, as part of the Wiley - La Trobe University agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Tanya Serry https://orcid.org/0000-0003-1538-7327
Tanya Stebbins https://orcid.org/0000-0002-9014-1944
Andrew Martchenko https://orcid.org/0000-0002-2618-9820
Natalie Araujo https://orcid.org/0000-0002-7555-8223
Brigid McCarthy https://orcid.org/0000-0002-1375-6726

ENDNOTES

* SMOG is a well-established measure in health communication whereby a lower score reflects a more readable text. SMOG is recommended by Cochrane: Cochrane Collaboration. Methodological Expectations of Cochrane Intervention Reviews (MECIR): Standards for the Reporting of Plain Language Summaries in New Cochrane Intervention Reviews (PLEACS). Chichester (UK): Wiley & Sons; 2013.

† Flesch Reading Ease is also widely used readability measure in which a higher score reflects greater readability. This measure is mentioned in the Australian Government’s Digital Transformation Strategy.29

§ Box and Whisper plot interpretation:

- The median is shown by the horizontal line in the centre of the box. The first quartile is shown at the far left of the box and the third quartile is shown at the far right of the box. Fifty percent of the data fall in this range.

- The minimum number is shown at the far left of the range.

- The maximum score is shown at the far right of the box. Fifty percent of the data fall in this range.

- The first quartile is shown at the far left of the box and the third quartile is shown at the far right of the box. Outliers are calculated if multiplying the inter-quartile range by 1.5.

REFERENCES

1. Basch CH, Mohlman J, Hillyer GC, Garcia P. Public health communication in time of crisis: readability of on-line COVID-19 information. Disaster Med Public Health Prep. 2020;14(5):635–7.

2. Azios JH, Bellon-Harn M, Dockens AL, Manchaiah V. Quality and readability of English-language internet information for aphasia. Int J Speech Lang Pathol. 2019;21(1):1–9.

3. Papadakos JK, Charow RC, Papadakos CJ, Moody LJ, Giuliani ME. Evaluating cancer patient-reported outcome measures: readability and implications for clinical use. Cancer. 2019;125(8):1350–6.

4. Szmuda T, Özdemir C, Ali S, Singh A, Syed MT, Stlonieowski P. Readability of online patient education material for the novel coronavirus disease (COVID-19): a cross-sectional health literacy study. Public Health. 2020;185:21–5.

5. Ferguson C, Merga M, Winn S. Communications in the time of a pandemic: the readability of documents for public consumption. Aust N Z J Public Health. 2021;45(2):116–21.

6. Tanton R, Peel D, Vidyattama Y. Every suburb Every town: poverty in Victoria. Victoria, Australia; 2018.

7. Victoria’s diverse population: 2016 Census State of Victoria, Victoria. 2017. Available from: https://www.vic.gov.au/sites/default/files/2019-08/Victorias-Diverse-Population-brochure-2016.pdf.

8. Mac OA, Muscat D, Ayre J, Patel P, McCaffery KJ. Coronavirus (COVID-19) vaccination information must pay attention to health literacy: analysis of readability of official COVID-19 public health information. Med J Aust. 2021;215:373–5.

9. Mac OA, Muscat DM, Ayre J, Patel P, McCaffery KJ. The readability of official public health information on COVID-19. Med J Aust. 2021; 215(8):373–5.

10. Smith JA, Judd J. COVID-19: vulnerability and the power of privilege in a pandemic. Health Promot J Austr. 2020;31(2):158–60.

11. Albright J, de Guzman C, Acoo P, Paiva D, Faulkner M, Swanson J. Readability of patient education materials: implications for clinical practice. Appl Nurs Res. 1996;9(3):139–43.

12. Boulos MNK. British internet-derived patient information on diabetes mellitus: is it readable? Diabetes Technol Ther. 2005;7(3):528–35.

13. James N. The readability debate 2016. Available from: https://www.plainenglishfoundation.com/documents/10179/54323/PEF-Neil_James-The_readability_debate.pdf.

14. Literacy and access. 2021. Available from: https://www.stylemanual.gov.au/user-needs/understanding-needs/literacy-and-access.

15. Programme for the International Assessment of Adult Competencies, Australia 2013. Available from: https://www.abs.gov.au/ausstats/abs@.nsf/DOSSbyTopic/0997C269D99F3EDBCA257B260079BE3E?OpenDocument.

16. McEllan T, Dobinson G. Tertiary education of journalists and the readability of Australian newspapers. Aust Journal Rev. 2004;26(1):155–74.

17. COVID-19: a chronology of state and territory government announcements (up until 30 June 2020) 2020 Oct 22. Available from: https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/rp2021/Chronologies/COVID-19StateTerritoryGovernmentAnnouncements.
18. Mishra V, Dexter JP. Comparison of readability of official public health information about COVID-19 on websites of international agencies and the governments of 15 countries. JAMA Netw Open. 2020;3(8):e2018033.

19. Wrigley Kelly N, Murray K, McCarthy C, O’Shea D. An objective analysis of quality and readability of online information on COVID-19. Health Technol. 2021;11(5):1093–9.

20. Worrall AP, Connelly MJ, O’Neill A, O’Doherty M, Thornton KP, McNally C, et al. Online readability of COVID-19 health information: a comparison between four English speaking countries. BMC Public Health. 2020;20(1):1635.

21. Halboub E, Al-Ak’hali MS, Al-Mekhlafi HM, Alhajj MN. Quality and readability of web-based Arabic health information on COVID-19: an infodemiological study. BMC Public Health. 2021;21(1):1–7.

22. Industrial-strength natural language processing in Python. 2021. Available from: https://spacy.io/.

23. Partalidou E, Spyromitros-Xioufis E, Doropoulos S, Vologiannidis S, Diamantaras KI. Design and implementation of an open source Greek POS tagger and entity recognizer using spaCy. 2019 IEEE/WIC/ACM international conference on web intelligence (WI); 2019 14–17 Oct; 2019.

24. Hüning M, Hüning M. TextStat simple text analysis tool. Berlin: Dutch Linguistics, Free University of Berlin; 2005.

25. Mc Laughlin GH. SMOG grading—a new readability formula. J Read. 1969;12(8):639–46.

26. Henry SL. Referencing and linking to WAI guidelines and technical documents. 2018. Available from: https://www.w3.org/WAI/standards-guidelines/linking/

27. Flesch R. How to write plain English. University of Canterbury. 1979 [cited 2016 Feb 5]. Available from: https://pages.stern.nyu.edu/~wstarbuc/Writing/Flesch.htm

28. Flesch R. A new readability yardstick. J Appl Psychol. 1948;32(3):221–33.

29. Clear language and writing style. Available from: https://www.stylemanual.gov.au/writing-and-designing-content/clear-language-and-writing-style#readability

30. Flesch RF. The Art of readable writing. 1949.

31. Berg SH, O’Hara JK, Shortt MT, Thune H, Brännick KK, Lungu DA, et al. Health authorities’ health risk communication with the public during pandemics: a rapid scoping review. BMC Public Health. 2021;21(1):1–23.

How to cite this article: Serry T, Stebbins T, Martchenko A, Araujo N, McCarthy B. Improving access to COVID-19 information by ensuring the readability of government websites. Health Promot J Austral. 2022. https://doi.org/10.1002/hpja.610