Readability of English-Language Self-Report Measures in Bipolar Disorder

Joseph Nickel, MD; Sujeeth Shanmugam, BS; Niki Jona, BS; Kaushal Shah, MD, MPH; and Predrag Gligorovic, MD, MHA

Abstract

Objective: Self-report measures have been used as a tool to evaluate mania and depressive symptoms and aid in diagnosing bipolar disorder. The American Medical Association recommends that the evaluation reading be at a fifth- to sixth-grade level. The research literature has previously established that questionnaires regarding other psychiatric conditions have been written at an elevated reading level compared to the recommended guidelines. The objective of this study was to calculate the overall readability of measures in bipolar disorder and the individual readability of the items and instructions sections.

Methods: Measures in assessing bipolar disorder were selected from a publicly available literature search in the PsycINFO, PubMed, PubMed Central, and Google Scholar databases. Nine English-language measures met the inclusion criteria and were analyzed. The readability of each measure was analyzed via 4 validated formulas, Gunning Fog, Simple Measure of Gobbledygook, FORCAST, and Flesch Reading Ease, through an external website (readable.com). Measures with an average readability greater than 6.00 were considered above the recommended reading level.

Results: All measures had at least 1 component (instructions or items) written above the recommended reading

level. The mean reading level of the instructions and items sections were 9.35 (SD = 1.45; range, 7.03–11.51) and 8.78 (SD = 1.73; range, 5.59–11.31), respectively.

Conclusion: The results indicate that using these measures in populations with low reading literacy may be a limitation in correctly identifying manic and depressive symptomology. If the self-report symptom tool is utilized with a higher readability than the average patient, it may not be guaranteed that this tool will improve the diagnosis process.

Prim Care Companion CNS Disord 2025;27(3):24m03892

*Author affiliations are listed at the end of this article.*

The utility of self-report questionnaires is limited by the comprehension abilities of their target audience. When written at an appropriate reading level, they can provide valuable insight to physicians in diagnosing patients with mental disorders such as bipolar disorder. When written at an elevated reading level, they may perpetuate health disparities within socioeconomic populations. The American Medical Association recommends that the reading level for such materials remains below the fifth- to sixth-grade level.[1] Research has shown that people with mental health challenges have reduced literacy compared to the general population.[2] Evidence suggests that individuals with mental illness may have limited available opportunities for reading and writing due to potential attentive and functional impairment, as well as reduced access to general education and vocational training.[2]

Research indicates that materials related to psychiatric conditions often exceed the recommended

reading levels for the public. A study on patient health pamphlets in mental health reported an average readability at the eighth-grade reading level.[3] Previous readability studies have found elevated reading levels of questionnaires in depression, anxiety, and binge-eating disorders.[4,5] It was reported that publicly available web-based information on bipolar disorder is currently written at an elevated reading level and often requires higher-level education to be utilized appropriately.[6]

Emotion and mood can strongly influence attention, perception, memory, and reasoning, which are crucial for reading comprehension.[7] Studies show that bipolar disorder mood swings or episodes may affect working memory processes.[8] Researchers also noted that symptoms of cognitive impairment remained during the resolution of the acute episode and euthymic periods of bipolar disorder.[9] Sleep deprivation may further exacerbate attention problems and impair memory, making it more challenging for individuals with bipolar

Scan Now
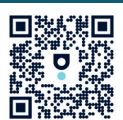
Cite and Share this article at Psychiatrist.com

Clinical Points

- Self-assessment measures are vital tools to help in the diagnosis of mental disorders including bipolar disorder.
- Readability of self-assessment measures should be at or below a sixth-grade reading level to be suitable for the general population.
- Currently, all assessed self-assessment measures had at least 1 component written above the sixth-grade reading level.

disorder to process information.[10] Evidence suggests that cognition is more impaired following partial compared to total sleep deprivation, further suggesting that these specific types of sleep restriction be given special attention within the context of bipolar disorder.[11] In addition, the first episode of bipolar disorder was associated with cognitive deficits that were widespread on neuropsychological testing.[12]

Pharmacologic treatments can have effects on cognition in bipolar disorder. In a study on second-generation antipsychotics in patients with bipolar disorder, risperidone- and olanzapine-treated patients performed poorly on working memory, attention, and verbal fluency tasks compared with healthy individuals.[13] In contrast, patients with bipolar disorder taking anticonvulsants or lithium had an improvement in cognitive performance.[14] A cross-sectional study comparing psychomotor speed, memory, attention, and executive functioning in bipolar disorder patients taking valproic acid, lamotrigine, carbamazepine, or lithium showed that lamotrigine-treated patients had the highest cognitive performance, followed by patients treated with lithium, valproic acid, and carbamazepine.[14]

Many have sought to develop a variety of screening tools that would aid in identifying people with bipolar disorder. While there are numerous available tools, their quality is of interest. This is the first study, to our knowledge, to analyze the readability of self-report measures in bipolar disorder. The objective of this study was to calculate the overall readability of measures in bipolar disorder and the individual readability of the items and instructions sections of measures in bipolar disorder. We hypothesize that the reading levels of these questionnaires will be elevated compared to the recommended standard.

METHODS

Systematic reviews and measures were selected from a publicly available literature search in the PsycINFO, PubMed, PubMed Central, and Google Scholar databases. We performed the literature search using the Medical Subject Heading term "bipolar disorders" combined with "surveys and questionnaires" to identify measures.

For all measures reviewed, the inclusions criteria were (1) self-reported, (2) written in the English language, (3) assessed symptoms of bipolar disorder, (4) publicly available, and (5) used for the assessment of adults. Exclusion criteria included clinician-observed measures, clinician standardized interviews, and measures not developed in English in the United States. The last date a database search was conducted was December 28, 2023.

Twenty-eight measures in bipolar disorder were identified from a systematic review by Cerimele and colleagues[15] in 2019 and a literature search, 19 of which were later excluded because they did not meet the above inclusion criteria or were not obtainable. Four of these 9 measures did not include standard instructions and thus were excluded from analyses involving instructions. The following measures were included: Altman Self-Rating Mania Scale, Carroll Depression Scale, Inventory of Depressive Symptoms Self-Report, Quick Inventory of Depressive Symptoms Self-Report, Affective Self-Rating Scale, Daily Mood Monitoring, Internal State Scale (ISS), Multidimensional Assessment of Thymic States (MAThyS), and NIMH Prospective Life-Chart Methodology–Self.[16–24]

Readability was assessed via readable software (readable.com). Each measure was analyzed via 4 readability formulas from the readable software: Gunning Fog, Simple Measure of Gobbledygook (SMOG), FORCAST, and Flesch Reading Ease.

The Gunning Fog score is computed with the following equation[25–27]:

$$grade\ level = 0.4 \times ((average\ sentence\ length) + \%\ polysyllable\ words)$$

The SMOG score is computed with the following equation[25–27]:

$$grade\ level = 3 + square\ root\ (polysyllable\ count)$$

The FORCAST score is computed with the following equation[25–27]:

$$grade\ level = 20 - (monosyllable\ words/10)$$

The Flesch Reading Ease score is computed with the following equation[28]:

$$score = 206.835 - 1.015 \times \left(\frac{words}{sentence} \right) - 84.6 \times \left(\frac{syllables}{words} \right)$$

These validated indices were included as they are commonly used throughout the United States to analyze text and have been previously used to assess the readability of measures and patient education materials in prior studies in psychiatry. These readability formulas utilize various forms of information to calculate each score, including the number of syllables per word, the length of sentences, and the degree

of deviation from basic word lists. The Gunning Fog, SMOG, and FORCAST indices provided a corresponding grade level (ie, sixth grade and seventh grade) based on the measure analyzed. However, the Flesch Reading Ease Score was calculated by a score and was manually converted to a grade level by the authors for analysis. The readability for each measure was calculated by averaging scores across all 4 indices. The readability of the item and instruction sections was analyzed separately. Measures with an average readability greater than 6.00 were considered above the recommended reading level. Mean and SD readability scores for each metric across measures were computed. Analysis of the data was performed using Microsoft Excel (version 16.49).

Table 1 illustrates the conversion metrics between the score and grade level for the Flesch Reading Ease Score index and highlights the differences between the readability indices.[29]

RESULTS

All measures had at least 1 component (instructions or items) written above the recommended reading level. All measures with instructions had a mean readability score above the recommended sixth-grade level. The items sections of 8 out of 9 measures included in our study had a mean readability score above the recommended sixth-grade level. The mean reading level of the instructions and items sections was 9.35 (SD = 1.45; range, 7.03–11.51) and 8.78 (SD = 1.73; range, 5.59–11.31), respectively.

A reading comprehension level of high school or above was required in 6 out of 9 measures (67%).

Figure 1 depicts the average readability of the instructions and items sections for each bipolar measure.[30] The readability analysis of various bipolar disorder assessment measures revealed a wide range of scores. The MAThyS measure exhibited the highest readability score, with instructions averaging 11.50 and items averaging 10.01, indicating a higher level of complexity. In contrast, the ISS measure had lower readability scores, with instructions averaging 7.03 and items averaging 5.59, suggesting it is easier to read and comprehend. This suggests that some measures may need to be modified to ensure they are accessible to readers across a broader spectrum of literacy.

Figure 2 highlights the readability scores of each measure using Gunning Fog, SMOG, FORCAST, and Flesch Reading Ease Score indices.[30] The readability of measures for bipolar disorder was evaluated using 4 indices: Flesch Reading Ease, Gunning Fog, SMOG, and FORCAST. The mean of the Gunning Fog index was 6.79, the Flesch Reading Ease index was 9.00, the SMOG index was 9.10, and the FORCAST was 10.22. The readability scores computed by the different indices correspond to reading grade levels in the United States.

DISCUSSION

The difficulties with the diagnosis of bipolar disorder have been discussed for decades from the physician's point

Table 1.

Differences and Scoring of the Readability Indices

Readability indices	Reading level calculation difference
Gunning Fog	The Gunning Fog is a weighted average of the number of words per sentence and the number of long words per word. The grade level corresponds to the following education level: • 6–8th grade: middle school • 9–12th grade: high school • 13–16th grade: college
SMOG	The SMOG formula scores a text based on the complexity of its sentences and words. Specifically, it considers the number of polysyllabic words (words with 3 or more syllables) and the number of sentences in a text
FORCAST	The grade-level calculation is based on the syllables in the 150-word sample. This focus makes it best for texts that do not have full sentence stops or the usual paragraph format
Flesch Reading Ease	The Flesch Reading Ease scores a text based on the number of syllables per word and the number of words per sentence. The index provides a numeric score with values between 0 and 100 corresponding with specific grade levels based on the following bulleted information: • 100–90: 5th grade • 90–80: 6th grade • 80–70: 7th grade • 70–60: 8th grade • 60–50: 11th grade • 50–30: college level • 30–10: college graduate • 10–0: professional

Figure 1.

Readability of Instructions and Items Sections of Measures in Bipolar Disorder[30]
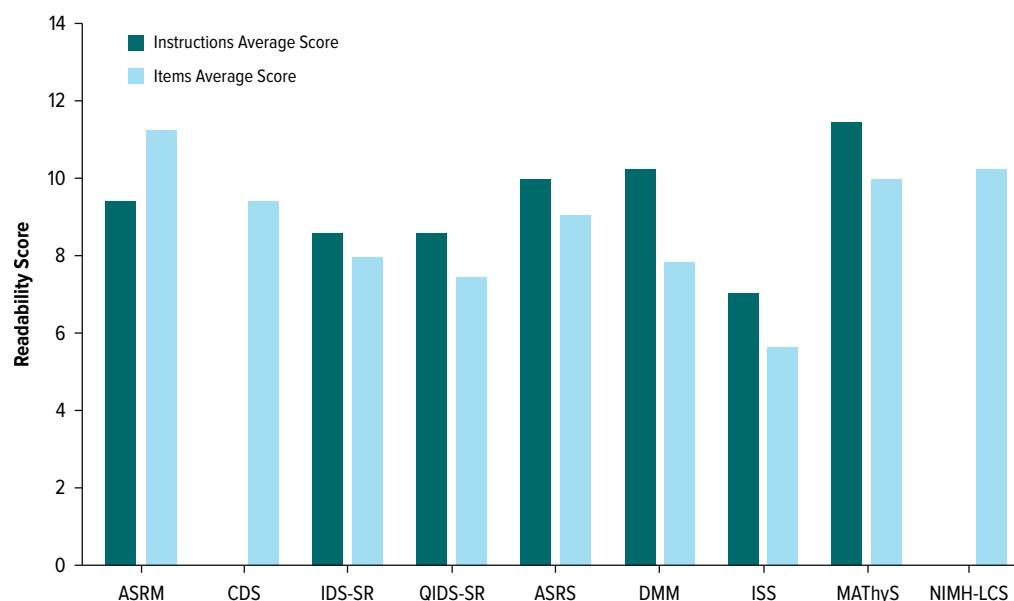

Abbreviations: ASRM = Altman Self-Rating Mania Scale, ASRS = Affective Self-Rating Scale, CDS = Carrol Depression Scale, DMM = Daily Mood Monitoring, IDS-SR = Inventory of Depressive Symptoms Self-Report, ISS = Internal State Scale, MAThyS = Multidimensional Assessment of Thymic States, NIMH-LCS = NIMH Prospective Life-Chart Methodology–Self, QIDS-SR = Quick Inventory of Depressive Symptoms Self-Report.

of view due to its purely clinical presentation and challenges with the identification of symptoms. Overdiagnosis of bipolar disorder was initially noted in the literature in 2007–2008.[31–33] Contributing factors for misdiagnosis were identified as personality disorders as possible confounders and patient history of substance abuse.[32] The *DSM-IV* was notable for introducing hypomania as a less severe subset of

symptoms than mania; however, the criteria for hypomania were less clearly defined, and diagnostic overlap between other psychiatric illnesses was not robustly incorporated in the diagnostic criteria, increasing the difficulty of making a correct diagnosis.

In 2009, the International Society for Bipolar Disorders recommended using symptom measurement

Figure 2.

Average Readability Score of Bipolar Measures Across Indices With Standard Deviation[30]
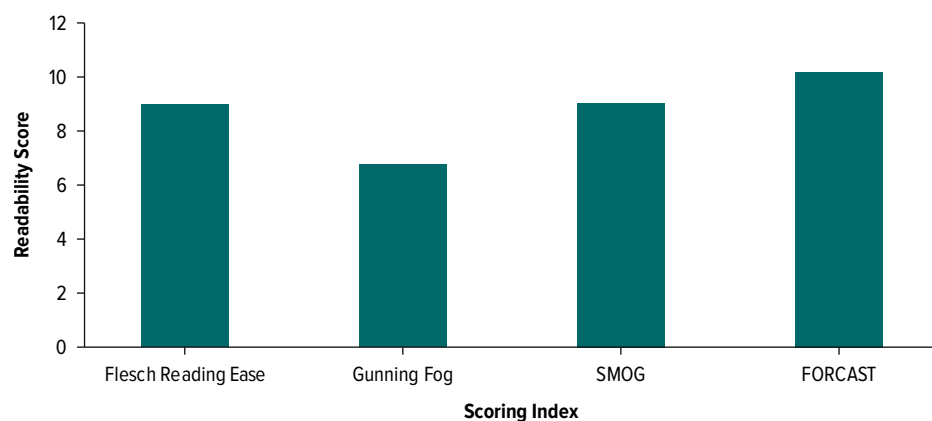
Abbreviation: SMOG = Simple Measure of Gobbledygook.

surveys in addition to history taking to diagnose bipolar disorder. After this, a study further supported their use in bipolar disorder with positive and negative predictive values greater than 85%.[34] When written at an appropriate reading level, these measures provide crucial information to physicians in the diagnosis of patients with mental disorders such as bipolar disorder. With the development of the *DSM-5* and self-report symptom screening tools, this dual approach to diagnosing patients with bipolar disorder has become a common standard. Even outside the health system, many young adults can find self-screening tools to assess their symptoms on the readily accessible internet.[35] Screening surveys tend to have high sensitivity but low specificity, potentially adding to the trend of overdiagnosis if the screening tool is utilized by itself.

Additionally, the accessibility and standardization of screening surveys to all populations have been called into question for many psychiatric disorders in the literature.[36] Doyen's[37] review critically analyzes strategies for specifically diagnosing bipolar disorder and discusses concerns regarding disproportionate misdiagnosis in communities with comorbid posttraumatic stress disorder and lower socioeconomic status. Thus, given the possibility of overdiagnosis or misdiagnosis, self-report questionnaires must be carefully scrutinized to limit potential downstream consequences amid the established difficulties with screening. With misdiagnosis comes unnecessary treatment with mood stabilizers and antipsychotics, and both classes of medications have significant and potentially dangerous side effects.

The empirical results reported should be considered in light of some limitations. Although readability formulas are valuable tools, they cannot capture the full complexity of medical jargon. Readability formulas do not consider qualitative factors such as layout, spacing, font, and text size, which may also impact readability. In this study, only the English versions were analyzed; therefore, these findings cannot be applied to any translations available in other languages. These readability formulas are based on grade levels in the United States and may not directly relate to reading levels of other English-speaking countries. The AMA and NIH guidelines also recommend patient-facing materials in the United States. They should not be generalized to other countries with different readability standards. For example, the UK National Health Services recommends that materials be written at or below a 9-year-old reading level.[38]

Our study focuses on evaluating the readability of screening surveys; with 9 out of 10 screening tools scoring above the average reading level in our study, the accuracy and utility of these measures to screen for mania and depression symptoms may not be standardized among each patient. A significant challenge in ensuring the effectiveness of the self-report questionnaire is ensuring that a patient is interpreting the questions appropriately. If the self-report symptom tool is utilized with a higher readability than the average patient, it may not be guaranteed that this tool will improve the diagnosis process. Amid the ongoing discussion of how best to diagnose bipolar disorder, our study results show that caution may be necessary before administering and utilizing this screening.

References

1. Health literacy: report of the council on scientific affairs. Ad Hoc committee on health literacy for the council on scientific affairs, American Medical Association. *JAMA*. 1999;281(6):552–557.
2. Sentell TL, Shumway MA. Low literacy and mental illness in a nationally representative sample. *J Nerv Ment Dis*. 2003;191(8):549–552.
3. Gralton E, Sher M, Lopez CD. Information and readability issues for psychiatric patients: e-learning for users. *Psychiatrist*. 2010;34(9):376–380.
4. McHugh RK, Behar E. Readability of self-report measures of depression and anxiety. *J Consult Clin Psychol*. 2009;77(6):1100–1112.
5. Richards LK, McHugh RK, Pratt EM, et al. Readability and comprehension of self-report binge eating measures. *Eat Behav*. 2013;14(2):167–170.
6. Piras M, Perra A, Gureje O, et al. The current quality of web-based information on the treatment of bipolar disorder: a systematic search. *J Clin Med*. 2022;11(18): 5427.
7. Tyng CM, Amin HU, Saad MNM, et al. The influences of emotion on learning and memory. *Front Psychol*. 2017;8:1454.
8. Soraggi-Frez C, Santos FH, Albuquerque PB, et al. Disentangling working memory functioning in mood states of bipolar disorder: a systematic review. *Front Psychol*. 2017;8:574.
9. Huang Y, Zhang Z, Lin S, et al. Cognitive impairment mechanism in patients with bipolar disorder. *Neuropsychiatr Dis Treat*. 2023;19:361–366.
10. Alhola P, Polo-Kantola P. Sleep deprivation: impact on cognitive performance. *Neuropsychiatr Dis Treat*. 2007;3(5):553–567.
11. Durmer JS, Dinges DF. Neurocognitive consequences of sleep deprivation. *Semin Neurol*. 2005;25(1):117–129.
12. Lee RSC, Hermens DF, Scott J, et al. A meta-analysis of neuropsychological functioning in first-episode bipolar disorders. *J Psychiatr Res*. 2014;57(1):1–11.
13. Torrent C, Martinez-Arán A, Daban C, et al. Effects of atypical antipsychotics on neurocognition in euthymic bipolar patients. *Compr Psychiatry*. 2011;52(6):613–622.
14. Gualtieri CT, Johnson LG. Comparative neurocognitive effects of 5 psychotropic anticonvulsants and lithium. *MedGenMed*. 2006;8(3):46.
15. Cerimele JM, Goldberg SB, Miller CJ, et al. Systematic review of symptom assessment measures for use in measurement-based care of bipolar disorders. *Psychiatr Serv*. 2019;70(5):396–408.
16. Altman EG, Hedeker D, Peterson JL, et al. The Altman self-rating mania scale. *Biol Psychiatry*. 1997;42(10):948–955.
17. Born C, Amann BL, Grunze H, et al. Saving time and money: a validation of the self ratings on the prospective NIMH life-chart method (NIMH-LCM). *BMC Psychiatry*. 2014;14(1):130.
18. Henry C, M'Bailara K, Mathieu F, et al. Construction and validation of a dimensional scale exploring mood disorders: MAThyS (Multidimensional Assessment of Thymic States). *BMC Psychiatry*. 2008;8(1):82.

19. Bauer MS, Vojta C, Kinosian B, et al. The Internal State Scale: replication of its discriminating abilities in a multisite, public sector sample. *Bipolar Disord*. 2000;2(4):340–346.

20. Schwartz S, Schultz S, Reider A, et al. Daily mood monitoring of symptoms using smartphones in bipolar disorder: a pilot study assessing the feasibility of ecological momentary assessment. *J Affect Disord*. 2016;191:88–93.

21. Adler M, Liberg B, Andersson S, et al. Development and validation of the Affective Self Rating Scale for manic, depressive, and mixed affective states. *Nord J Psychiatry*. 2008;62(2):130–135.

22. Brown ES, Murray M, Carmody TJ, et al. The Quick Inventory of Depressive Symptomatology-Self-report: a psychometric evaluation in patients with asthma and major depressive disorder. *Ann Allergy Asthma Immunol*. 2008;100(5):433–438.

23. Wardenaar KJ, van Veen T, Giltay EJ, et al. The structure and dimensionality of the Inventory of Depressive Symptomatology Self Report (IDS-SR) in patients with depressive disorders and healthy controls. *J Affect Disord*. 2010;125(1–3):146–154.

24. Merten T. [The Carroll Scale for Depression-the German version and validation]. *Psychiatr Neurol Med Psychol*. 1990;42(6):340–347.

25. Lee SE, Farzal Z, Kimple AJ, et al. Readability of patient-reported outcome measures for chronic rhinosinusitis and skull base diseases. *Laryngoscope*. 2020;130(10):2305–2310.

26. Dorismond C, Farzal Z, Thompson NJ, et al. Readability analysis of pediatric otolaryngology patient-reported outcome measures. *Int J Pediatr Otorhinolaryngol*. 2021;140:110550.

27. Lee SE, Farzal Z, Ebert CS, et al. Readability of patient-reported outcome measures for head and neck oncology. *Laryngoscope*. 2020;130(12):2839–2842.

28. Cherla D, Sanghvi S, Choudhry OJ, et al. Readability assessment of internet-based patient education materials related to endoscopic sinus surgery. *Laryngoscope*. 2012;122(8):1649–1654.

29. Rao SJ, Nickel JC, Kiell EP, et al. Readability of commonly used patient-reported outcome measures in laryngology. *Laryngoscope*. 2022;132(5):1069–1074.

30. Readability score | Readability test | Reading level calculator | Readable. Accessed August 6, 2024. https://readable.com/

31. Mitchell PB. Bipolar disorder: the shift to overdiagnosis. *Can J Psychiatry*. 2012;57(11):659–665. Accessed July 7, 2024. www.TheCJP.ca

32. Goldberg JF, Garno JL, Callahan AM, et al. Overdiagnosis of bipolar disorder among substance use disorder inpatients with mood instability. *J Clin Psychiatry*. 2008;69(11):1751–1757. Accessed July 7, 2024. https://www.psychiatrist.com/jcp/overdiagnosis-bipolar-disorder-among-substance-disorder

33. Ghouse AA, Sanches M, Zunta-Soares G, et al. Overdiagnosis of bipolar disorder: a critical analysis of the literature. *ScientificWorldJournal*. 2013;2013:297087.

34. Zimmerman M, Galione JN. Screening for bipolar disorder with the mood disorders questionnaire: a review. *Harv Rev Psychiatry*. 2011;19(5):219–228.

35. Kruzan KP, Meyerhoff J, Nguyen T, et al. "I wanted to see how bad it was": online self-screening as a critical transition point among young adults with common mental health conditions. *Proc SIGCHI Conf Hum Factor Comput Syst*. 2022;2022:328.

36. Green JG, Gruber MJ, Kessler RC, et al. Diagnostic validity across racial and ethnic groups in the assessment of adolescent DSM-IV disorders. *Int J Methods Psychiatr Res*. 2012;21(4):311–320.

37. Doyen P. The overdiagnosis of bipolar disorder within marginalized communities: a call to action. *Columbia Soc Work Rev*. 2021;19(1):7388.

38. *Content Design: Planning, Writing and Managing Content - Writing for GOV.UK - Guidance - GOV*. UK. Accessed August 8, 2024. https://www.gov.uk/guidance/content-design/writing-for-gov-uk