OPEN

PRS GLOBAL OPEN

ORIGINAL ARTICLE
Education

Readability of Patient-reported Outcome Measures Used in Plastic Surgery

Zachary Zamore, BA*†
Chao Long Azad, MD, MPH*†
Lily Zhu, BS†
Chenery Lowe, PhD, ScM‡§
Aviram M. Giladi, MD, MS*

Background: Patient-reported outcomes are essential to understanding success in plastic surgery procedures, many that aim to improve quality of life. Patient-reported outcome measures (PROMs) should be written at or below the sixth-grade reading level recommended by the American Medical Association. This study aimed to evaluate the readability of plastic surgery PROMs.

Methods: We conducted a literature review to identify validated, commonly used PROMs in plastic surgery. We extracted PROMs' text and instructions and analyzed readability using different approaches that estimate the grade level required to understand. Our primary outcome was the Simple Measure of Gobbledygook (SMOG) index, which detects word complexity and expects 100% comprehension at the grade level rating assigned. We also included the Flesch-Kincaid grade level, Coleman-Liau index, and automated readability index.

Results: Forty-three PROMs met the inclusion criteria. The mean SMOG index was 8.2 (SD = 1.3), indicating an eighth-grade reading level. Mean reading grade levels measured by the Flesch-Kincaid grade level, Coleman-Liau index, and automated readability index ranged from third to sixth grade, although these may underestimate readability difficulties. Only 6 (14%) PROMs had a SMOG index at or below the sixth-grade level. PROM instructions had significantly higher reading levels than the questions/responses for all readability indexes ($P < 0.01$).

Conclusions: PROMs used in plastic surgery, including the instructions, exceed the reading level recommended by the American Medical Association. This may limit comprehension and accurate completion and compromise validity and reliability. PROMs should be written and designed to be accessible to patients of all literacy levels. *(Plast Reconstr Surg Glob Open 2024; 12:e6351; doi: 10.1097/GOX.0000000000006351; Published online 20 December 2024.)*

INTRODUCTION

Patient-reported outcomes (PROs), collected using patient-reported outcome measures (PROMs), are critical to understanding patients' perspective of their health status.[1] PROs are particularly important for evaluating the success of plastic surgery procedures that often aim to improve quality of life and are not adequately captured by standard outcome measures such as mortality or complication rates. PROs are invaluable in assessing patients' views about their symptoms, day-to-day functioning, and overall quality of life.[2] In addition, PROs are used for research, quality monitoring, and improvement efforts.[2]

There are patient-, provider-, and clinic-level barriers to collecting PROs.[3,4] One patient-level barrier of particular concern for ensuring equitable care delivery, research, and quality assurance is patients' print literacy; patients cannot reliably self-administer PROMs if their print literacy is lower than the reading level at which PROMs are written.[5] In the United States, an estimated 54% of adults cannot read above the sixth-grade level.[6,7] Given that patients with low literacy have increased difficulty accessing healthcare and are at a higher risk of poor health,[8–11]

*From the *The Curtis National Hand Center, MedStar Union Memorial Hospital, Baltimore, MD; †Department of Plastic and Reconstructive Surgery, Johns Hopkins Medicine, Baltimore, MD; ‡Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; and §Stanford Center for Biomedical Ethics, Stanford University, Palo Alto, CA.*

Received for publication July 31, 2024; accepted October 9, 2024.

Presented at Plastic Surgery The Meeting, October 26–29, 2023, Austin, TX.

DOI: 10.1097/GOX.0000000000006351

Disclosure statements are at the end of this article, following the correspondence information.

Related Digital Media are available in the full-text version of the article on www.PRSGlobalOpen.com.

failure to capture these patients' perspectives via PROMs may perpetuate and exacerbate these known disparities.

The American Medical Association (AMA) recommends patient-facing materials be written at or below the sixth-grade reading level.[12] Previous studies have examined the readability of plastic surgery-related patient resources and educational materials and have shown that most fail to meet the AMA guidelines for readability.[13–20] In other surgical specialties, recent studies have also shown that PROM compliance with these recommendations varies.[21–26] Only 12% of orthopedic PROMs met the AMA readability guidelines of sixth-grade reading level or lower, and no PROMs used in anesthesia met this readability target.[21,26] Previous studies also found that many PROM instructions did not meet AMA guidelines for readability.[24,27] This study aimed to evaluate the readability of the questions/answers and patient instructions of plastic surgery–specific PROMs.

Takeaways

Question: Are plastic surgery-related patient-reported outcome measures (PROMs) written at the sixth-grade reading level or lower, as recommended by the American Medical Association for patient-facing materials?

Findings: We extracted the text from plastic surgery PROMs and analyzed readability. We found that on average, these PROMs are written at the eighth-grade reading level. The readability of plastic surgery PROMs has not improved over time.

Meaning: Low and limited literacy plastic surgery patients may have difficulty understanding and accurately completing PROMs, which may impede outcome evaluations, research, and quality improvement efforts. PROMs should be written to be accessible to patients of all literacy levels.

METHODS

PROM Selection

We conducted a literature review to identify commonly used and validated plastic surgery PROMs. Initial references were extracted from Dobbs et al[28] and Sharma et al[2] that included lists of PROMs used in various plastic surgery subspecialties. We supplemented this list with PubMed searches for subspecialty-specific systematic reviews for PROMs used in hand and upper extremity,[29] burn,[30] gender affirmation,[31,32] craniofacial,[33,34] peripheral nerve,[35,36] lower extremity,[37] and breast.[38] All searches were conducted between January 2023 and May 2024 using keywords "Plastic surgery" and/or "subspecialty" and "Patient-reported outcome measures" or "PROM." The inclusion criteria included PROMs that were designed for disease processes that plastic surgeons treat, validated in a plastic surgery patient population, and available in English. "Validated" was defined as PROMs that underwent both qualitative evaluation and quantitative/psychometric testing. In cases where multiple versions of a PROM existed, the most recent version was used for analysis.

Readability Measures and Analysis

We extracted the text content of each PROM and its associated instructions for analysis. We also collected the year the PROM questionnaire was published, the country in which it was developed, and the total number of items in the PROM. We selected the Simple Measure of Gobbledygook (SMOG) index, which measures word complexity, as the primary outcome measure to assess readability. Our secondary outcome measures included the Flesch–Kincaid grade level (FK), the Coleman–Liau index (CLI), and the automated readability index (ARI). The FK is used most frequently in the literature and assesses readability based on sentence length and syllables per word.[39,40] The ARI and CLI both evaluate readability based on sentence length and character count per word.[41,42] All readability measurements are used to assign a US grade level required to understand the text. Readability formulas are outlined in Figure 1.

We chose the SMOG index as our primary outcome measure due to its consistency, widespread usage in healthcare settings, and simple methodology.[39] In addition, it is based on a 100% comprehension expectation, meaning that it assigns the lowest grade level in which 100% comprehension can be expected.[43] For example, if a text has a SMOG score of 8, an eighth grader would be expected

$$SMOG = 1.0430\sqrt{Number\ of\ polysyllabic\ words \times \frac{30}{Number\ of\ sentences}}$$

$$FK = 0.39\left(\frac{Total\ words}{Total\ sentences}\right) + 11.8\left(\frac{Total\ syllables}{Total\ words}\right) - 15.59$$

$$CLI = \left(0.0588 \times \left(\frac{Average\ \#\ of\ letters}{Per\ 100\ words}\right)\right) - \left(0.296 \times \left(\frac{Average\ \#\ of\ sentences}{Per\ 100\ words}\right)\right) - 15.8$$

$$ARI = 4.71\left(\frac{Total\ characters}{Total\ words}\right) + 0.5\left(\frac{Total\ words}{Total\ sentences}\right) - 21.43$$

Fig. 1. Formulas used to calculate the readability measures.

to comprehend 100% of the text.[43] This contrasts with the FK and CLI, which assign the lowest grade level at which approximately 75% comprehension is expected.[40,44] The ARI assigns grade levels based on schools' reference texts with grade-level determinations made with inconsistent methods[42,45] and likely not based on 100% comprehension, as many reading materials in schools are assigned a grade level designed to challenge the reader to improve comprehension skills.[45] Because PROMs are designed, developed, and validated with the assumption or goal that patients understand all or almost all of the instruments, a readability metric based on 100% comprehension expectation is most appropriate.[46] Furthermore, although 75% comprehension might be acceptable for most reading materials, it is reasonable to demand a higher level of comprehension for health-related materials that often include information that is important for the patient to fully understand.[39]

We used Readable.com to conduct the analysis. PROM text was prepared following US Department of Health and Human Services guidelines to avoid inflated readability scores.[47] For example, fill-in-the-blank questions with multiple answer choices were replaced with full sentences for each answer choice, and embedded formatting such as bullet points were removed.[48]

Statistical Analysis

We calculated descriptive statistics, including mean and SD, for all readability measures. We used the Shapiro–Wilk test to evaluate the normality of our data. We conducted independent *t* tests to compare the readability measures between the PROMs' questions/responses versus the instructions. We conducted a correlation analysis between SMOG scores and the year of PROM questionnaire publication, as well as between SMOG scores and the number of items in the PROM instrument, by calculating the Pearson correlation coefficient. To assess the robustness of these findings, we conducted a sensitivity analysis by removing outliers (defined as >2 SDs from the mean) and evaluated how this impacted the Pearson coefficient. We used Microsoft Excel 2016 and Python 3.11 software for the analysis. Significance was defined as a *P* value less than 0.05.

RESULTS

We identified 87 PROMs, 43 of which met our inclusion criteria (Fig. 2). (**See table, Supplemental Digital Content 1,** which displays the list of PROMs included for analysis and their readability scores [GBR = the United Kingdom, USA = the United States of America, SWE = Sweden, AUS = Australia, CAN = Canada, NLD = the Netherlands, NOR = Norway, BEL = Belgium, DNK = Denmark, ITA = Italy], **http://links.lww.com/PRSGO/D659**.[49–86]) The mean SMOG index was 8.2 (SD = 1.3), indicating a mean reading level above the eighth grade. Only 6 (14%) PROMs had a SMOG index of sixth grade or lower: CLEFT-Q, FACE-Q, the European Organisation for Research and Treatment of Cancer Head and Neck Module (EORTC QLQ-H&N43), Nasal Obstruction Symptom Evaluation Scale, Sheffield Pelvic Organ Prolapse Quality of Life Questionnaire, and the Derriford Appearance Scale (DAS59) (**Supplemental Digital Content 1, http://links.lww.com/PRSGO/D659**). For secondary readability measures, mean FK = 4.7 (SD = 1.6), CLI = 6.0 (SD = 2.2), and ARI = 3.3 (SD = 1.7), indicating reading levels between the third and sixth grades.

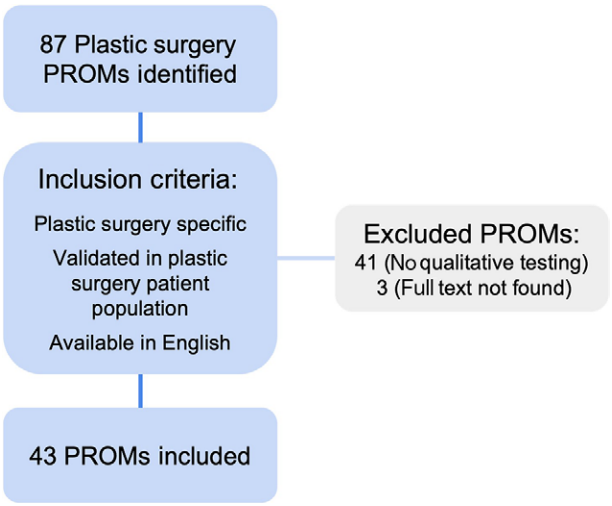

Fig. 2. PROM inclusion flow chart.

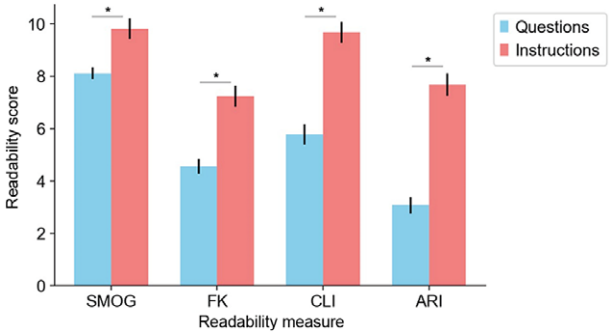

Fig. 3. Average readability score of PROM questions/responses and instructions. *$P < 0.01$.

When comparing PROM instructions to the questions/responses, PROM instructions had a significantly higher reading level for all readability indexes ($P < 0.01$ for SMOG, FK, CLI, and ARI) (Fig. 3). The average SMOG score for instructions was 9.8 (SD = 2.4), compared with 8.1 (SD = 1.3) for the questions/answers. Only 4 (11.8%) of the 34 PROMs with instructions had instructions with a SMOG index at or below the sixth-grade level.

The average number of questions per PROM was 431 (SD = 53.4). There was a negative correlation between readability and the number of items in the PROM ($r = -0.38$, $P = 0.01$) (Fig. 4). After removing outliers (FACE-Q and BODY-Q) in the sensitivity analysis, there was still a negative correlation between SMOG index and the number of items in the PROM ($r = -0.33$, $P = 0.04$). There was no significant correlation between the SMOG readability score and the year of PROM questionnaire publication (range, 1987–2020) ($r = -0.004$, $P = 0.98$).

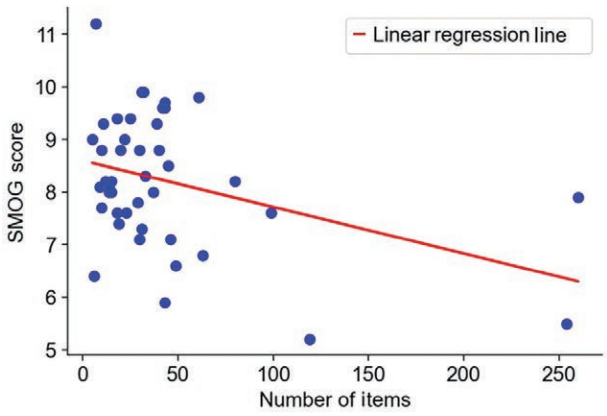

Fig. 4. Correlation between SMOG index and the number of items in a PROM. $r = -0.39$, $P = 0.01$.

DISCUSSION

Based on the SMOG metric, we found that only 14% of PROMs validated for use in plastic surgery patient populations met the AMA readability guidelines. These PROMs are written on average at the eighth-grade reading level, whereas the instructions are written at the 10th-grade reading level. Patients with low and limited literacy may have difficulty understanding and completing PROMs accurately, which may compromise PROMs' validity and reliability.[46] Patients with low literacy of plastic surgery may not be able to be evaluated comprehensively and may be excluded from research and quality improvement efforts.

We opted for the SMOG index as our primary outcome measure due to its foundation on a 100% comprehension expectation.[43] This contrasts with the FK and CLI, which assign the lowest grade level where a subject can achieve approximately 75% comprehension.[40,44,87] The ARI, on the other hand, assigns grade levels based on reference texts, relying on grade-level determinations established in 1967 by the Cincinnati School System and the publisher of the reference texts.[42] These determinations are arbitrary, inconsistent, and are likely not based on 100% comprehension given that many reading materials in schools are assigned a grade level with the goal of challenging the reader to enhance reading comprehension skills.[45] We deemed the 100% comprehension expectation established by the SMOG most appropriate for evaluating health-related materials such as PROMs.

We found that PROMs' instructions are written at an even higher reading level, as measured by all readability measurements. This may be because validation and evaluation processes for PROMs typically focus on the instrument items, whereas the instructions may be neglected or an afterthought.[88] Best practices in PROM development indicate that clarity of instructions should be explicitly evaluated during the development process.[46] Patients with low literacy who have difficulty reading PROMs' instructions may be more likely to complete PROMs incorrectly.[46]

We found that PROMs with more items tend to be more readable. This is possibly influenced by the fact that the 4 PROMs with the greatest number of questions, the BREAST-Q, FACE-Q, CLEFT-Q, and BODY-Q, have undergone a rigorous 3-phase development and validation process.[89] These trends persisted even after excluding outliers in our sensitivity analysis. Importantly, a large number of items increases the cognitive burden and survey fatigue, a known barrier to PROM completion.[90] Even though longer PROMs are correlated with greater readability, PROM length should still be minimized when possible.

We also found that there has been no improvement in the readability of PROMs over time: PROMs developed recently are no more readable than those developed decades ago, despite the well-known impacts of literacy on health and the policies that have followed (eg, AMA guidance on PROMs,[12] Plain Language Writing Act of 2010).[91] Moving forward, to better meet the needs of patients with low and limited literacy, PROM developers can improve the readability of both questions/answers and instructions by taking advantage of helpful resources such as automated readability assessment tools (such as Readable.com), artificial intelligence tools,[92] the Centers for Disease Control and Prevention Health Literacy website,[93] and PlainLanguage.gov.[94] PROM developers can also recruit patients with diverse literacy, educational, and linguistic backgrounds in the development and validation process, and this should be reported in studies to improve transparency. For already existing PROMs, these can be staff-administered to patients with low literacy, although this would increase cost and potentially cause patient embarrassment.[95] Alternatively, text-based PROMs can be adapted to a multimedia format that facilitates self-administration by patients of all literacy levels.[96]

This study has several limitations. First, none of the readability measures were developed to analyze health-related information and complex medical terminology. In addition, these measures detect various aspects of word and/or sentence complexity that may be imperfect proxies for readability.[97] However, the SMOG index is used often in healthcare settings and has several advantages, as previously detailed. Finally, because we did not perform a systematic review of plastic surgery PROMs, it is possible that our list of validated plastic surgery PROMs included for analysis is not comprehensive.

CONCLUSIONS

We found that most plastic surgery PROMs including the associated instructions are written above the recommended sixth-grade reading level. This may compromise accurate PRO measurement in plastic surgery, exacerbate existing health disparities, and limit the inclusion of patients with low literacy in research and quality improvement efforts. The readability of PROMs has not improved over time, suggesting that readability is an area needing increased attention in future PROM development efforts.

Aviram M. Giladi, MD, MS
The Curtis National Hand Center
MedStar Union Memorial Hospital
3333 N Calvert Street, JPB #200
Baltimore, MD 21218
E-mail: editor@curtishand.com
Twitter/X: cnhc_research; theaviram
Instagram: cnhc_research

DISCLOSURE

The authors have no financial interest to declare in relation to the content of this article.

REFERENCES

1. US Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. 2009. Available at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims. Accessed January 4, 2024.
2. Sharma K, Steele K, Birks M, et al. Patient-reported outcome measures in plastic surgery: an introduction and review of clinical applications. *Ann Plast Surg.* 2019;83:247–252.
3. Nguyen H, Butow P, Dhillon H, et al. A review of the barriers to using patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs) in routine cancer care. *J Med Radiat Sci.* 2021;68:186–195.
4. Yamamoto R, Choi C, Sayyed AA, et al. Challenges of large-scale patient-reported outcome measures collection in a multidisciplinary limb salvage center. *Plast Reconstr Surg Glob Open.* 2022;10:e4551.
5. Long C, Beres LK, Wu AW, et al. Patient-level barriers and facilitators to completion of patient-reported outcomes measures. *Qual Life Res.* 2022;31:1711–1718.
6. National Center for Education Statistics, US Department of Education. Program for the International Assessment of Adult Competencies (PIAAC). 2012/2014. Available at https://nces.ed.gov/surveys/piaac/national_results.asp. Accessed January 4, 2024.
7. Rothwell J. Assessing the economic gains of eradicating illiteracy nationally and regionally in the United States. 2020. Available at https://www.barbarabush.org/wp-content/uploads/2020/09/BBFoundation_GainsFromEradicatingIlliteracy_9_8.pdf. Accessed January 4, 2024.
8. Dewalt DA, Berkman ND, Sheridan S, et al. Literacy and health outcomes: a systematic review of the literature. *J Gen Intern Med.* 2004;19:1228–1239.
9. AD Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. 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;281:552–557.
10. Baker DW, Parker RM, Williams MV, et al. The health care experience of patients with low literacy. *Arch Fam Med.* 1996;5:329–334.
11. Weiss BD, Hart G, McGee DL, et al. Health status of illiterate adults: relation between literacy and health status among persons with low literacy skills. *J Am Board Fam Pract.* 1992;5:257–264.
12. Weiss B. Health Literacy: A Manual for Clinicians. American Medical Association Foundation and American Medical Association. 2003. Available at http://lib.ncfh.org/pdfs/6617.pdf. Accessed January 4, 2023.
13. Tran BNN, Singh M, Singhal D, et al. Readability, complexity, and suitability of online resources for mastectomy and lumpectomy. *J Surg Res.* 2017;212:214–221.
14. Seth AK, Vargas CR, Chuang DJ, et al. Readability assessment of patient information about lymphedema and its treatment. *Plast Reconstr Surg.* 2016;137:287e–295e.
15. Vargas CR, Ricci JA, Lee M, et al. The accessibility, readability, and quality of online resources for gender affirming surgery. *J Surg Res.* 2017;217:198–206.
16. Chen AD, Ruan QZ, Bucknor A, et al. Social media: is the message reaching the plastic surgery audience? *Plast Reconstr Surg.* 2019;144:773–781.
17. Vargas CR, Kantak NA, Chuang DJ, et al. Assessment of online patient materials for breast reconstruction. *J Surg Res.* 2015;199:280–286.
18. Fanning JE, Okamoto LA, Levine EC, et al. Content and readability of online recommendations for breast implant size selection. *Plast Reconstr Surg Glob Open.* 2023;11:e4787.
19. Ricci JA, Vargas CR, Chuang DJ, et al. Readability assessment of online patient resources for breast augmentation surgery. *Plast Reconstr Surg.* 2015;135:1573–1579.
20. Patel AA, Joshi C, Varghese J, et al. Do websites serve our patients well? A comparative analysis of online information on cosmetic injectables. *Plast Reconstr Surg.* 2022;149:655e–668e.
21. El-Daly I, Ibraheim H, Rajakulendran K, et al. Are patient-reported outcome measures in orthopaedics easily read by patients? *Clin Orthop Relat Res.* 2016;474:246–255.
22. Perez JL, Mosher ZA, Watson SL, et al. Readability of orthopaedic patient-reported outcome measures: is there a fundamental failure to communicate? *Clin Orthop Relat Res.* 2017;475:1936–1947.
23. Dorismond C, Farzal Z, Thompson NJ, et al. Readability analysis of pediatric otolaryngology patient-reported outcome measures. *Int J Pediatr Otorhinolaryngol.* 2021;140:110550.
24. Taylor DJ, Jones L, Edwards L, et al. Patient-reported outcome measures in ophthalmology: too difficult to read? *BMJ Open Ophthalmol.* 2021;6:e000693.
25. Stefu J, Slavych BK, Zraick RI. Patient-reported outcome measures in voice: an updated readability analysis. *J Voice.* 2023;37:465.e27–465.e34.
26. Nosow LM, Rao SJ, Neubauer DJ, et al. Readability of patient-reported outcome measures in anesthesiology. *Anesthesiology.* 2022;136:242–244.
27. Douglas A, Kelly-Campbell RJ. Readability of patient-reported outcome measures in adult audiologic rehabilitation. *Am J Audiol.* 2018;27:208–218.
28. Dobbs TD, Hughes S, Mowbray N, et al. How to decide which patient-reported outcome measure to use? A practical guide for plastic surgeons. *J Plast Reconstr Aesthet Surg.* 2018;71:957–966.
29. Marks M. Which patient-reported outcomes shall we use in hand surgery? *J Hand Surg Eur Vol.* 2020;45:5–11.
30. Griffiths C, Guest E, White P, et al. A systematic review of patient-reported outcome measures used in adult burn research. *J Burn Care Res.* 2017;38:e521–e545.
31. Andreasson M, Georgas K, Elander A, et al. Patient-reported outcome measures used in gender confirmation surgery: a systematic review. *Plast Reconstr Surg.* 2018;141:1026–1039.
32. Dy GW, Nolan IT, Hotaling J, et al. Patient reported outcome measures and quality of life assessment in genital gender confirming surgery. *Transl Androl Urol.* 2019;8:228–240.
33. Dobbs TD, Gibson JAG, Hughes S, et al. Patient-reported outcome measures for soft-tissue facial reconstruction: a systematic review and evaluation of the quality of their measurement properties. *Plast Reconstr Surg.* 2019;143:255–268.
34. Rhee JS, McMullin BT. Outcome measures in facial plastic surgery: patient-reported and clinical efficacy measures. *Arch Facial Plast Surg.* 2008;10:194–207.
35. Rayner MLD, Brown HL, Wilcox M, et al. Quantifying regeneration in patients following peripheral nerve injury. *J Plast Reconstr Aesthet Surg.* 2020;73:201–208.
36. de Moraes AA, Dantas DS, Chagas ACS, et al. Linking assessment instruments for brachial plexus injury to the international classification of functioning, disability and health. *J Hand Ther.* 2023;36:885–894.
37. Grigor EJM, Bitoiu B, Zeitouni C, et al. Patient-reported outcomes following free flap lower extremity reconstruction: a systematic review and meta-analysis. *J Plast Reconstr Aesthet Surg.* 2023;76:251–267.

38. Davies CF, Macefield R, Avery K, et al. Patient-reported outcome measures for post-mastectomy breast reconstruction: a systematic review of development and measurement properties. *Ann Surg Oncol*. 2021;28:386–404.
39. Wang LW, Miller MJ, Schmitt MR, et al. Assessing readability formula differences with written health information materials: application, results, and recommendations. *Res Social Adm Pharm*. 2013;9:503–516.
40. Kincaid JP, Fishburne RP, Jr, Rogers RL, et al. Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for navy enlisted personnel. Institute for Simulation and Training. 1975;56. Available at https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1055&context=istlibrary. Accessed January 4, 2024.
41. Coleman M, Liau TL. A computer readability formula designed for machine scoring. *J Appl Psychol*. 1975;60:283–284.
42. Smith EA, Senter RJ. Automated readability index. *AMRL TR*. 1967;60:1–14.
43. McLaughlin GH. SMOG grading—a new readability formula. *J Read*. 1969;12:639–646.
44. Flesch R. A new readability yardstick. *J Appl Psychol*. 1948;32:221–233.
45. Mills RE, Richardson JR, Richardson JE. What do publishers mean by "grade level?". *Read Teach*. 1963;16:359–362.
46. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2—assessing respondent understanding. *Value Health*. 2011;14:978–988.
47. Agency for Healthcare Research and Quality. Tip 6. Use caution with readability formulas for quality reports. 2015. Available at https://www.ahrq.gov/talkingquality/resources/writing/tip6.html. Accessed January 4, 2024.
48. Mac O, Ayre J, Bell K, et al. Comparison of readability scores for written health information across formulas using automated vs manual measures. *JAMA Netw Open*. 2022;5:e2246051.
49. Klassen AF, Cano SJ, Scott A, et al. Assessing outcomes in body contouring. *Clin Plast Surg*. 2014;41:645–654.
50. Levine DW, Simmons BP, Koris MJ, et al. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. *J Bone Joint Surg Am*. 1993;75:1585–1592.
51. Temple CL, Cook EF, Ross DC, et al. Development of a Breast Reconstruction Satisfaction Questionnaire (BRECON): dimensionality and clinical importance of breast symptoms, donor site issues, patient expectations, and relationships. *J Surg Oncol*. 2010;101:209–216.
52. Sigurdson L, Kirkland SA, Mykhalovskiy E. Validation of a questionnaire for measuring morbidity in breast hypertrophy. *Plast Reconstr Surg*. 2007;120:1108–1114.
53. Pusic AL, Klassen AF, Scott AM, et al. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plast Reconstr Surg*. 2009;124:345–353.
54. Waljee JF, Kim HM, Burns PB, et al. Development of a brief, 12-item version of the Michigan Hand Questionnaire. *Plast Reconstr Surg*. 2011;128:208–220.
55. Munster AM, Horowitz GL, Tudahl LA. The abbreviated Burn-Specific Health Scale. *J Trauma*. 1987;27:425–428.
56. Kildal M, Andersson G, Fugl-Meyer AR, et al. Development of a brief version of the Burn Specific Health Scale (BSHS-B). *J Trauma*. 2001;51:740–746.
57. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)—a simple practical measure for routine clinical use. *Clin Exp Dermatol*. 1994;19:210–216.
58. Carr T, Moss T, Harris D. The DAS24: a short form of the Derriford Appearance Scale DAS59 to measure individual responses to living with problems of appearance. *Br J Health Psychol*. 2005;10:285–298.
59. Harris DL, Carr AT. The Derriford Appearance Scale (DAS59): a new psychometric scale for the evaluation of patients with disfigurements and aesthetic problems of appearance. *Br J Plast Surg*. 2001;54:216–222.
60. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind Med*. 1996;29:602–608.
61. Thomson HJ, Winters ZE, Brandberg Y, et al. The early development phases of a European Organisation for Research and Treatment of Cancer (EORTC) module to assess patient reported outcomes (PROs) in women undergoing breast reconstruction. *Eur J Cancer*. 2013;49:1018–1026.
62. Bjordal K, Ahlner-Elmqvist M, Tollesson E, et al. Development of a European Organization for Research and Treatment of Cancer (EORTC) questionnaire module to be used in quality of life assessments in head and neck cancer patients. EORTC Quality of Life Study Group. *Acta Oncol*. 1994;33:879–885.
63. Klassen AF, Cano SJ, Scott A, et al. Measuring patient-reported outcomes in facial aesthetic patients: development of the FACE-Q. *Facial Plast Surg*. 2010;26:303–309.
64. Kahn JB, Gliklich RE, Boyev KP, et al. Validation of a patient-graded instrument for facial nerve paralysis: the FaCE scale. *Laryngoscope*. 2001;111:387–398.
65. VanSwearingen JM, Brach JS. The Facial Disability Index: reliability and validity of a disability assessment instrument for disorders of the facial neuromuscular system. *Phys Ther*. 1996;76:1288–1298; discussion 1298.
66. Herbenick D, Reece M. Development and validation of the female genital self-image scale. *J Sex Med*. 2010;7:1822–1830.
67. Martin RL, Irrgang JJ, Burdett RG, et al. Evidence of validity for the Foot and Ankle Ability Measure (FAAM). *Foot Ankle Int*. 2005;26:968–983.
68. Robinson K, Gatehouse S, Browning GG. Measuring patient benefit from otorhinolaryngological surgery and therapy. *Ann Otol Rhinol Laryngol*. 1996;105:415–422.
69. Mancuso CA, Lee SK, Saltzman EB, et al. Development of a questionnaire to measure impact and outcomes of brachial plexus injury. *J Bone Joint Surg Am*. 2018;100:e14.
70. Binkley JM, Stratford PW, Lott SA, et al. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network. *Phys Ther*. 1999;79:371–383.
71. Davis AM, Wright JG, Williams JI, et al. Development of a measure of physical function for patients with bone and soft tissue sarcoma. *Qual Life Res*. 1996;5:508–516.
72. Gabel CP, Melloh M, Burkett B, et al. Lower limb functional index: development and clinimetric properties. *Phys Ther*. 2012;92:98–110.
73. Chung KC, Pillsbury MS, Walters MR, et al. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. *J Hand Surg Am*. 1998;23:575–587.
74. Moolenburgh SE, Mureau MA, Duivenvoorden HJ, et al. Validation of a questionnaire assessing patient's aesthetic and functional outcome after nasal reconstruction: the patient NAFEQ-score. *J Plast Reconstr Aesthet Surg*. 2009;62:656–662.
75. Stewart MG, Witsell DL, Smith TL, et al. Development and validation of the Nasal Obstruction Symptom Evaluation (NOSE) scale. *Otolaryngol Head Neck Surg*. 2004;130:157–163.
76. Durani P, McGrouther DA, Ferguson MW. The Patient Scar Assessment Questionnaire: a reliable and valid patient-reported outcomes measure for linear scars. *Plast Reconstr Surg*. 2009;123:1481–1489.

77. Cano SJ, Browne JP, Lamping DL, et al. The patient outcomes of surgery-hand/arm (POS-hand/arm): a new patient-based outcome measure. *J Hand Surg Br.* 2004;29:477–485.
78. Cano SJ, Browne JP, Lamping DL, et al. The patient outcomes of surgery-head/neck (POS-head/neck): a new patient-based outcome measure. *J Plast Reconstr Aesthet Surg.* 2006;59:65–73.
79. Hays RD, Spritzer KL, Amtmann D, et al. Upper-extremity and mobility subdomains from the Patient-Reported Outcomes Measurement Information System (PROMIS) adult physical functioning item bank. *Arch Phys Med Rehabil.* 2013;94:2291–2296.
80. Beaton DE, Wright JG, Katz JN; Upper Extremity Collaborative Group. Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am.* 2005;87:1038–1046.
81. Spector IP, Carey MP, Steinberg L. The sexual desire inventory: development, factor structure, and evidence of reliability. *J Sex Marital Ther.* 1996;22:175–190.
82. Bradshaw HD, Hiller L, Farkas AG, et al. Development and psychometric testing of a symptom index for pelvic organ prolapse. *J Obstet Gynaecol.* 2006;26:241–252.
83. Rhee JS, Matthews BA, Neuburg M, et al. Creation of a quality of life instrument for nonmelanoma skin cancer patients. *Laryngoscope.* 2005;115:1178–1185.
84. Chren MM, Lasek RJ, Quinn LM, et al. Skindex, a quality-of-life measure for patients with skin disease: reliability, validity, and responsiveness. *J Invest Dermatol.* 1996;107:707–713.
85. Hill B, Pallant J, Williams G, et al. Evaluation of internal construct validity and unidimensionality of the brachial assessment tool, a patient-reported outcome measure for brachial plexus injury. *Arch Phys Med Rehabil.* 2016;97:2146–2156.
86. McGuire JK, Berg D, Catalpa JM, et al. Utrecht Gender Dysphoria Scale–Gender Spectrum (UGDS-GS): construct validity among transgender, nonbinary, and LGBQ samples. *Int J Transgend Health.* 2020;21:194–208.
87. Rankin EF, Culhane JW. Comparable cloze and multiple-choice comprehension test scores. *J Read.* 1969;13:193–198.
88. Morley D, Dummett S, Kelly L, et al. The Oxford Participation and Activities Questionnaire: study protocol. *Patient Relat Outcome Meas.* 2013;5:1–6.
89. Wong Riff KW, Tsangaris E, Goodacre T, et al. International multiphase mixed methods study protocol to develop a cross-cultural patient-reported outcome instrument for children and young adults with cleft lip and/or palate (CLEFT-Q). *BMJ Open.* 2017;7:e015467.
90. Sokas C, Hu F, Edelen M, et al. A review of PROM implementation in surgical practice. *Ann Surg.* 2022;275:85–90.
91. CONGRESS.GOV. H.R.946—Plain Language Act of 2010. 111th Congress (2009–2010); 2010.
92. Ali R, Connolly ID, Tang OY, et al. Bridging the literacy gap for surgical consents: an AI-human expert collaborative approach. *medRxiv.* 2023.
93. Centers for Disease Control and Prevention. Health literacy. Available at https://www.cdc.gov/healthliteracy/index.html. Accessed December 29, 2023.
94. plainlanguage.gov. The plain language action and information network. US General Services Administration. Available at https://www.plainlanguage.gov/. Accessed December 29, 2023.
95. Jahagirdar D, Kroll T, Ritchie K, et al. Using patient reported outcome measures in health services: a qualitative study on including people with low literacy skills and learning disabilities. *BMC Health Serv Res.* 2012;12:431.
96. Long C, Beres LK, Wu AW, et al. Developing a protocol for adapting multimedia patient-reported outcomes measures for low literacy patients. *PLoS One.* 2021;16:e0252684.
97. Redish JC, Selzer J. The place of readability formulas in technical communication. *Tech Commun.* 1985;32:46–52.