

Readability of Orthopaedic Patient-reported Outcome Measures: Is There a Fundamental Failure to Communicate?

Jorge L. Perez MD, Zachary A. Mosher BS, Shawna L. Watson BA, Evan D. Sheppard MD, Eugene W. Brabston MD, Gerald McGwin Jr. MS, PhD, Brent A. Ponce MD

Received: 28 December 2016 / Accepted: 27 March 2017 / Published online: 3 April 2017
© The Association of Bone and Joint Surgeons® 2017

Abstract

Background Patient-reported outcome measures (PROMs) are increasingly used to quantify patients' perceptions of functional ability. The American Medical Association and NIH suggest patient materials be written at or below 6th to 8th grade reading levels, respectively, yet one recent study asserts that few PROMs comply with these recommendations, and suggests that the majority of PROMs are written at too high of a reading level for self-administered patient use. Notably, this study was limited in its use of only one readability algorithm, although there is no commonly accepted, standard readability algorithm for healthcare-related materials. Our study, using multiple readability equations and heeding equal weight to each, hopes to yield a broader, all-encompassing estimate of readability, thereby offering a more accurate assessment of the readability of orthopaedic PROMS.

Questions/Purposes (1) What proportion of orthopaedic-related PROMs and orthopaedic-related portions of the

NIH Patient Reported Outcomes Measurement Information System (PROMIS®) are written at or below the 6th and 8th grade levels? (2) Is there a correlation between the number of questions in the PROM and reading level? (3) Using systematic edits based on guidelines from the Centers for Medicare and Medicaid Services, what proportion of PROMs achieved American Medical Association and NIH-recommended reading levels?

Methods Eighty-six (86) independent, orthopaedic and general wellness PROMs, drawn from commonly referenced orthopaedic websites and prior studies, were chosen for analysis. Additionally, owing to their increasing use in orthopaedics, four relevant short forms, and 11 adult, physical health question banks from the PROMIS®, were included for analysis. All documents were analyzed for reading grade levels using 19 unique readability algorithms. Descriptive statistics were performed using SPSS Version 22.0.

Results The majority of the independent PROMs (64 of 86; 74%) were written at or below the 6th grade level, with 81 of 86 (94%) written at or below the 8th grade level. All item banks (11 of 11) and short forms (four of four) of the PROMIS® were written below the 6th grade reading level. The median reading grade level of the 86 independent PROMs was 5.0 (interquartile range [IQR], 4.6–6.1). The PROMIS® question banks had a median reading grade level of 4.1 (IQR, 3.5–4.8); the Adult Short Forms had a median reading grade level of 4.2 (IQR, 4.2–4.3) There was no correlation appreciated between the median reading grade level and the number of questions contained in a PROM ($r = -0.081$; $p = 0.460$). For PROMs above NIH-recommended levels, following edits, all (five of five) achieved NIH reading level goals and three (three of five) achieved American Medical Association goals. Editing of these PROMs improved readability by 4.3 median grade

Each author certifies that he or she has no commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connections with the submitted article.

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research*® editors and board members are on file with the publication and can be viewed on request.

Electronic supplementary material The online version of this article (doi:10.1007/s11999-017-5339-0) contains supplementary material, which is available to authorized users.

J. L. Perez, Z. A. Mosher, S. L. Watson, E. D. Sheppard, E. W. Brabston, G. McGwin Jr., B. A. Ponce (✉)
Division of Orthopaedic Surgery, Department of Surgery,
University of Alabama at Birmingham, 1313 13th Street South,
Birmingham, AL 35205, USA
e-mail: bponce@uabmc.edu

level (before, 8.9 [IQR, 8.4–9.1], after 4.6 [IQR, 4.6–6.4], difference of medians, 4.3; $p = 0.008$).

Conclusions Patient literacy has great influence on healthcare outcomes, and readability is an important consideration in all patient-directed written materials. Our study found that more than 70% of PROMs commonly used in orthopaedics, and all orthopaedic-related portions of the PROMIS[®] are written at or below the most stringent recommendations (\leq 6th grade reading level), and more than 90% of independent PROMs and all PROMIS[®] materials are written at or below an 8th grade level. Additionally, the use of the Centers for Medicare and Medicaid Services guidelines for editing high reading level PROMs yields satisfactory results.

Clinical Relevance Fears of widely incomprehensible PROMs may be unfounded. Future research to identify the most appropriate readability algorithm for use in the healthcare sector, and revalidation of PROMs after readability-improving edits is warranted.

Introduction

The CDC defines “health literacy” as “the degree to which an individual has the capacity to obtain, communicate, process, and understand basic health information” [127]. Although correlation does not equate to causation, previous studies have shown associations between health literacy and healthcare outcomes, noting that poor health literacy correlates with increased healthcare cost, hospitalization rates, and mortality [35, 95, 118, 119]. Accordingly, the CDC encourages physicians to maximize health literacy by adapting patient materials to the level of knowledge of the intended audience [99]. For written materials, the American Medical Association (AMA) suggests writing at or below the 6th grade reading level; similarly, the NIH suggests writing at the 7th or 8th grade reading level [128, 135]. This “readability” can be determined by using various validated formulas which incorporate factors such as sentence length, word count, and word complexity. While studies have shown that patient-directed materials, such as consent forms, educational materials, or discharge summaries often are written above suggested reading levels [7, 24, 123, 135], few studies have analyzed materials used by patients to self-report their health status [3, 38]. Known as “patient-reported outcome measures,” (PROMs) these questionnaires are used clinically and in research to quantify patients’ perceptions of their conditions, functional abilities, baseline health status, treatment success, and physician competency [6, 8, 18, 103, 108].

Validation of PROMs is important to ensure that they measure the endpoints of interest accurately and

reproducibly. Elements such as reliability, consistency, content/construct validity, and sensitivity to change, are often considered part of this validation process; however, readability is seldom mentioned as a factor in validation [11, 88, 97, 110]. Thus, while a PROM may have elements of a well-designed survey, low readability could impair its practical value and clinical utility. However, readability is not synonymous with comprehension, as comprehension is a multifaceted concept of aesthetic utility and academic content. Thus, while patient materials may be deemed “highly readable,” if they are written in poor font or color choices, the reader may have an issue comprehending the contained content. Furthermore, with diverse knowledge levels among patients, the level of comprehension is unique to every patient, regardless of the readability level of the PROM content. Even so, investigating the readability of patient materials offers practitioners a sense for whether broad, diverse populations of patients are likely to be able to use these tools in real-world practice.

Two studies exist regarding the readability of orthopaedic-specific PROMs [3, 38], and both are limited in their scope of PROM selection and their heavy reliance on a single readability measurement test, the Flesch Reading Ease [3, 38]. While the Flesch Reading Ease is one of the most commonly-used readability scores in health literature, its continued utility with modern syntax has been called into question as newer, more broadly applicable readability measures have been developed [130]. In the absence of a single, accepted, validated readability measure for healthcare materials, the use of a lone readability measure could lead to an unnecessary skew in the results of these studies. To account for this, one systematic review supports using multiple readability measures to evaluate a passage [45]. Additionally, prior studies have not assessed if PROMs with higher numbers of questions were traditionally written at higher reading levels. With trends toward increasingly brief surveys [37], questions arise regarding if lack of readability was an issue in longer surveys. Finally, if PROMs were written at too high of a reading level, an effort must be made to improve their readability and warrant their continued clinical use. While this has been shown in patient education materials [52, 115], this has not been performed in PROMs. Therefore, questions arise regarding if the same techniques used to improve patient education materials are also applicable for use with PROMs, and if edited versions of PROMs continue to possess their prior reliability and validation.

We therefore asked: (1) What proportion of orthopaedic-related PROMs and orthopaedic-related portions of the NIH Patient Reported Outcomes Measurement Information System (PROMIS[®]) are written at or below the 6th and 8th grade levels? (2) Is there a correlation between the number of questions in the PROM and reading level? (3) Using

systematic edits based on guidelines from Centers for Medicare and Medicaid Services (CMS) [90], what proportion of PROMs achieved NIH-recommended reading levels?

Materials and Methods

Selection of Patient-Reported Outcomes Instrument List

A PubMed search was conducted to identify an inclusive list of orthopaedic-associated PROMs. The most relevant article identified, “Are patient-reported outcome measures in orthopaedics easily read by patients?” included a list of 59 PROMs [38]. We supplemented this list with orthopaedic PROMs from the following resources: (1) “Guide to outcomes instruments for musculoskeletal trauma research,” if they were specified as “patient” reported, and not “combined” or “physician” reported [1]; (2) the American Academy of Orthopaedic Surgeons’ (AAOS) website [5]; and (3) the quick-link Orthopaedic Scores website [74].

Preparation of Patient-Reported Outcomes Documents

A total of 86 independent PROMs were identified for inclusion and obtained in their published form (ie, as original journal publications or via the authors’ respective

websites). These PROMs were grouped as follows: general health/musculoskeletal/pain status (15) (Table 1), upper extremity (21) (Table 2), lower extremity (41) (Table 3), and spine (nine) (Table 4). In addition, four PROMIS® Adult Short Forms and one investigator-compiled “PROMIS® Bank” consisting of questions from 11 relevant PROMIS® Adult Item Banks were assessed (Table 5). Individual PROMs and PROMIS® materials were attained in Portable Document Format (PDF), manually converted to Microsoft Word® format (Microsoft Corporation, Redmond, WA, USA), and reviewed for accuracy by the authors. All advertisements, hyperlinks, pictures, copyright notices, and other text that was not a direct element of the questionnaire were removed. Each PROM or section of the PROMIS® (item bank or short form) then was saved as a text-only file for analysis by the readability software.

Readability Assessment

Readability tests were chosen based on the following inclusion criteria: (1) intended for English text; (2) intended for adult use or used in a previously published study; and (3) score output scale of grade level, with higher grade levels corresponding to a more difficult to comprehend text. Additionally, we included the Flesch Reading Ease readability index score (scale, 1–100) owing to its simple grade scale convertibility and for comparative relevance with previously published use [38]. In the absence of any, single accepted readability measure for healthcare-related

Table 1. Median reading grade levels of 15 common, orthopaedic-related, patient-reported outcome measures for general/musculoskeletal health or pain status, as determined by 19 unique readability algorithms

Patient-reported outcome measure	MGL (25th and 75th percentiles)	IQR	Number of questions
General Health/Musculoskeletal/Pain Status	4.9 (4.3, 5.0)	0.7	
Short Form-McGill Pain Questionnaire [93]	6.7 (5.6, 10.1)	4.5	18
EQ-5D Questionnaire [39]	6.1 (5.3, 8.3)	3.0	6
EQ-5D (United Kingdom Version) [40]	5.1 (4.8, 7.8)	3.0	6
Stanford Health Assessment Questionnaire (20-item) [46]	5.0 (4.6, 7.4)	2.8	20
Dallas Pain Questionnaire [76]	5.0 (4.6, 6.4)	1.8	16
Rheumatoid and Arthritis Outcome Score [19]	5.0 (4.1, 6.2)	2.1	42
Physical Activity Scale for the Elderly [133]	4.9 (4.1, 7.0)	2.9	12
Short Musculoskeletal Function Assessment [122]	4.9 (3.2, 6.1)	2.9	46
Stanford Health Assessment Questionnaire (8-item) [106]	4.8 (3.8, 7.2)	3.4	8
McGill Pain Questionnaire [92]	4.6 (3.8, 6.3)	2.5	21
Roland Morris Disability Questionnaire [111]	4.5 (4.2, 5.8)	1.6	24
Veterans RAND 12 (SF-12) [131]	4.1 (2.5, 5.1)	2.6	12
Veterans RAND 36 (SF-36) [132]	4.0 (2.0, 4.8)	2.8	36
Veterans RAND 20 (SF-20) [121]	3.9 (2.1, 4.5)	2.4	20
Nottingham Health Profile [55]	2.6 (0.2, 3.8)	3.6	38

MGL = Median grade level; IQR = interquartile range; EQ-5D = EuroQol Five Dimensions.

Table 2. Median reading grade levels of 21 common, orthopaedic-related, patient-reported outcome measures of the upper extremity, as determined by 19 unique readability algorithms

Patient-reported outcome measure	MGL (25th and 75th percentiles)	IQR	Number of questions
Upper extremity	5.1 (4.6, 5.7)	1.1	
Western Ontario Arthritis of the Shoulder Index [79]	7.7 (5.4, 9.0)	3.6	19
DASH with Sports/Performing Arts and Work Modules [53]	7.6 (5.8, 9.4)	3.6	38
Western Ontario Shoulder Instability Index [69]	7.0 (5.7, 9.1)	3.4	21
Western Ontario Rotator Cuff Index [68]	7.0 (5.3, 8.9)	3.6	20
QuickDASH with Sports/Performing Arts and Work Modules [9]	6.5 (5.3, 8.3)	3.0	19
Kerlan-Jobe Orthopaedic Clinic Shoulder and Elbow Score [2]	5.7 (4.8, 7.3)	2.5	10
Shoulder Pain and Disability Index [110]	5.6 (4.4, 8.4)	4.0	13
Mayo Wrist Score [4]	5.4 (5.0, 7.2)	2.2	5
Michigan Hand Outcomes Questionnaire [25]	5.4 (4.2, 6.7)	2.5	37
Shoulder Rating Questionnaire [75]	5.2 (4.4, 8.1)	3.7	21
Boston Carpal Tunnel Questionnaire [78]	5.1 (4.7, 7.2)	2.5	19
Constant Shoulder Score [27]	5.1 (4.4, 7.3)	2.9	6
Mayo Elbow Performance Score [96]	5.0 (4.2, 7.3)	3.1	3
Oxford Shoulder Score [31]	4.9 (4.2, 7.3)	3.1	12
Subjective Shoulder Rating System [71]	4.6 (3.5, 5.9)	2.4	6
Oxford Elbow Score [30]	4.6 (3.0, 6.3)	3.3	12
Oxford Shoulder Instability Score [32]	4.4 (2.9, 5.4)	2.5	12
Patient-Rated Wrist Evaluation [83]	4.4 (2.5, 5.1)	2.6	15
ASES-Elbow Score [67]	4.3 (1.8, 6.1)	4.3	19
Simple Shoulder Test [87]	3.9 (2.3, 4.7)	2.4	14
ASES-Shoulder Score [109]	3.8 (0.9, 5.2)	4.1	17

MGL = median grade level; IQR = interquartile range; ASES = American Shoulder and Elbow Surgeons.

materials, each document was analyzed by 19 unique readability algorithms, each meeting the criteria above (Table 6). Assessment was performed via Readability Studio 2015 (Oleander Software, Ltd, Pune, Maharashtra, India). Descriptions and algorithms for each readability test were adapted from the Readability Studio descriptions (Appendix 1. Supplemental material is available with the online version of *CORR*[®]).

Descriptive Statistics

Descriptive statistics were performed on the readability test results, and the median grade level (MGL) and interquartile range (IQR) were reported. Spearman's correlation coefficient was used to determine whether the number of survey items in each PROM correlated with its readability level. All statistical analyses were performed using SPSS Version 22.0 (IBM SPSS Statistics for Macintosh, Armonk, NY, USA).

Readability Improvement Editing Process

For PROMs with mean readability scores above the 8th grade level, the following editing steps, based on the CMS

Toolkit for Making Written Material Clear and Effective [90], were instituted. We edited the PROMs by using active voice, simple, short sentences, and a simplified vocabulary [90]. After these three steps, the median grade level (MGL) was reassessed. All PROMs meeting criteria for inclusion underwent each editing step as outlined above (Appendix 2. Supplemental material is available with the online version of *CORR*[®]).

Results

Sixty-four of 86 PROMs (74%) were found to have an MGL at or below the AMA-recommended 6th grade reading level, while 81 of 86 of the scores (94%) were found to be at or below the NIH-recommended 8th grade level (Fig. 1). The overall MGL of independent PROMs was 5.0 (IQR, 4.6–6.1), corresponding to approximately the start of the United States' 5th grade school year. The investigator-compiled PROMIS[®] Bank had an MGL of 4.1 (IQR, 3.5–4.8). The four selected PROMIS[®] Adult Short Forms had an MGL of 4.2 (IQR, 4.2–4.3) (Table 5). The Nottingham Health Profile has the lowest MGL of the independent PROMS (MGL, 2.6; IQR, 0.2–3.8) (Table 1), followed by the American Shoulder and Elbow Surgeons Questionnaire

Table 3. Median reading grade levels of 41 common, orthopaedic-related, patient-reported outcome measures of the lower extremity, as determined by 19 unique readability algorithms

Patient-reported outcome measure	MGL (25th and 75th percentiles)	IQR	Number of questions
Lower extremity	5.2 (4.6, 6.6)	2.0	
UCLA Activity Score [136]	12.1 (7.0, 13.9)	6.9	10
Modified Cincinnati Rating System [102]	9.1 (6.2, 10.5)	4.3	8
Lower Extremity Measure [59]	8.9 (5.5, 10.9)	5.4	29
Lysholm Knee Score [82]	8.4 (6.7, 13.2)	6.5	8
Tegner Activity Level Scale [124]	8.4 (6.1, 9.8)	3.7	5
Kujala Questionnaire [73]	7.7 (5.7, 9.5)	3.8	13
Foot and Ankle Ability Measure [85]	7.3 (5.6, 9.1)	3.5	22
Foot and Ankle Disability Index-Sport [49]	7.0 (5.6, 9.1)	3.5	8
Majeed Pelvic Score [84]	6.8 (5.4, 8.4)	4.0	7
Foot and Ankle Disability Index [49]	6.7 (5.5, 8.9)	3.4	26
Lower Extremity Functional Scale [13]	6.6 (5.5, 9.5)	4.0	20
International Knee Documentation Committee Score [57]	6.4 (5.4, 8.8)	3.4	19
Knee Disability and Osteoarthritis Outcome Score-Physical Function Short Form [104]	6.0 (5.2, 7.2)	2.0	7
Lower Extremity Activity Scale [114]	5.9 (5.2, 7.5)	2.3	12
Knee Outcome Survey Sports Activities Scale [58]	5.7 (5.1, 8.3)	3.2	11
Hip Disability and Osteoarthritis Outcome Score-Physical Function Short Form [29]	5.7 (5.1, 7.8)	2.7	5
Foot and Ankle Outcome Score [112]	5.7 (4.7, 7.0)	2.3	42
Western Ontario Meniscal Evaluation Tool [70]	5.5 (4.9, 7.0)	2.1	16
HOOS [101]	5.3 (4.4, 6.6)	2.2	40
KOOS [113]	5.2 (4.3, 6.5)	2.2	42
WOMAC of the Hip [11]	5.2 (4.0, 6.5)	2.5	24
Copenhagen Hip and Groin Outcome Score [126]	5.1 (4.8, 6.9)	2.1	37
VAS for Calcaneal Fractures [51]	5.0 (4.9, 7.7)	2.8	13
Knee Outcome Survey Activities of Daily Living Scale [58]	5.0 (4.6, 7.4)	2.8	14
KOOS - Joint Replacement [80]	5.0 (4.5, 6.8)	2.3	7
HOOS - Joint Replacement [81]	5.0 (4.5, 6.7)	2.2	6
AAOS-Lower Limb Questionnaire [61]	5.0 (4.4, 6.9)	2.5	7
Knee Society Clinical Rating System/Knee Society Score [56]	5.0 (4.4, 5.8)	1.4	66
Index of Severity for Osteoarthritis of the Knee (Lequesne) [77]	4.8 (3.8, 6.2)	2.4	11
AAOS-Foot and Ankle Module [61]	4.7 (3.8, 6.3)	2.5	25
Index of Severity for Osteoarthritis of the Hip (Lequesne) [77]	4.6 (3.1, 6.1)	3.0	11
Kaikkonen Functional Scale [65]	4.6 (2.7, 6.5)	3.8	9
Forgotten Joint Score [10]	4.6 (2.5, 5.7)	3.2	12
Oxford Knee Score [34]	4.5 (3.0, 5.5)	2.5	12
Oxford Hip Score [33]	4.5 (3.0, 5.4)	2.4	12
Foot Health Status Questionnaire [12]	4.5 (3.0, 5.1)	2.1	13
Hip Rating Questionnaire [60]	4.5 (2.9, 5.7)	2.8	14
Ankle Osteoarthritis Scale [36]	4.5 (2.1, 6.5)	4.4	18
Revised Foot Function Index [21]	4.4 (2.7, 5.6)	2.9	68
Iowa Pelvic Score [125]	4.4 (2.3, 6.9)	4.6	25
Marx Activity Rating Scale [86]	3.9 (2.0, 4.5)	2.5	4

MGL = median grade level; IQR = interquartile range; UCLA = University of California, Los Angeles; KOOS = Knee Disability and Osteoarthritis Outcome Score; HOOS = Hip Disability and Osteoarthritis Outcome Score; AAOS = American Academy of Orthopaedic Surgeons.

Table 4. Median reading grade levels of nine common, orthopaedic-related, patient-reported outcome measures of the spine, as determined by 19 unique readability algorithms

Patient-reported outcome measure	MGL (25th and 75th percentiles)	IQR	Number of questions
Spine	5.0 (4.8, 6.6)	1.8	
Bournemouth Back Questionnaire [15]	8.0 (5.4, 8.5)	3.1	7
Bournemouth Neck Questionnaire [16]	8.0 (5.4, 8.5)	3.1	7
Quebec Back Pain Disability Scale [72]	6.6 (5.4, 8.4)	3.0	20
Neck Disability Index [129]	5.0 (4.4, 6.6)	2.2	10
Copenhagen Neck Functional Disability Scale [63]	5.0 (3.5, 6.6)	3.1	15
Revised Oswestry Disability Questionnaire [54]	4.9 (4.4, 6.4)	2.0	10
Oswestry Disability Index (Version 1) [41]	4.8 (4.2, 6.1)	1.9	10
Oswestry Disability Index (Version 2) [42]	4.8 (4.1, 5.5)	1.4	10
Neck Outcome Score [64]	4.6 (3.4, 6.2)	2.8	34

MGL = median grade level; IQR = interquartile range.

Table 5. Median reading grade levels of the NIH PROMIS[®] question sets

NIH PROMIS [®] Question Set	MGL (25th and 75th percentiles)	IQR
Physical Health Question Bank	4.1 (3.5, 4.8)	1.3
Upper Extremity	5.1 (3.6, 7.4)	3.8
Physical Function with Mobility	5.0 (3.4, 7.3)	3.9
Physical Function	4.8 (2.9, 6.8)	3.9
Physical Function-Cancer	4.8 (2.8, 6.6)	3.8
Mobility	4.6 (2.5, 6.6)	4.1
Fatigue-Cancer	3.8 (1.8, 5.2)	3.4
Pain Interference-Cancer	4.1 (1.8, 4.8)	3.0
Pain Interference	4.0 (1.8, 4.7)	2.9
Fatigue	3.2 (1.2, 4.9)	3.7
Sleep-related Impairment	2.4 (1.2, 4.4)	3.2
Pain Behavior	2.3 (1.2, 4.3)	3.1
Questionnaires	4.2 (4.2, 4.3)	0.1
PROMIS [®] -10	4.6 (3.2, 6.3)	3.1
PROMIS [®] -29	4.2 (1.9, 5.3)	3.4
PROMIS [®] -43	4.2 (1.9, 5.2)	3.3
PROMIS [®] -57	4.2 (1.8, 5.0)	3.2

MGL = median grade level; NIH PROMIS[®] = NIH Patient-reported Outcomes Measurement Information System; IQR = interquartile range.

(MGL, 3.8; IQR, 0.9–5.2) (Table 2), Marx Activity Rating Scale (MGL, 3.9; IQR, 2.0–4.5) (Table 3), RAND 20-item Short Form (MGL, 3.9; IQR, 2.1–4.5) (Table 1), and Simple Shoulder Test (MGL, 3.9; IQR, 2.3–4.7) (Table 2). The PROMs with the highest MGLs were the UCLA Activity Score (MGL, 12.1; IQR, 7.0–13.9), Modified Cincinnati Rating System (MGL, 9.1; IQR, 6.2–10.5), Lower Extremity Measure (MGL, 8.9; IQR, 5.5–10.9), Lysholm Knee Score (MGL, 8.4; IQR, 6.7–13.2), and the Tegner Activity Level Scale (MGL, 8.4; IQR, 6.1–9.8) (Table 3). All item banks and short forms of the PROMIS[®] achieved AMA and NIH recommendations (Table 5).

There was no correlation appreciated between the MGL and the number of questions contained in a PROM ($r = -0.081$; $p = 0.460$).

Following edits, all five PROMs (UCLA Activity Score, Modified Cincinnati Rating System, Lysholm Knee Score, Tegner Activity Level Scale, Lower Extremity Measure) achieved the NIH-recommended 8th grade level, while three (Modified Cincinnati Rating System, Tegner Activity Level Scale, Lower Extremity Measure) achieved the AMA recommendation of 6th grade level (Fig. 2). Editing of these PROMs improved readability by 4.3 MGL (before: 8.9 [IQR, 8.4–9.1], after: 4.6 [IQR 4.6–6.4]; difference of median, 4.3; $p = 0.008$).

Table 6. Readability tests with MGL and IQR

Readability test	MGL (25th and 75th percentiles)	IQR
FORCAST [22]	10.7 (9.6, 11.7)	2.1
Bormuth Grade Placement [17]	8.4 (7.9, 9.0)	1.1
Flesch Kincaid Reading Ease [43]	8.2 (7.5, 10.1)	2.6
SMOG [91]	8.0 (7.3, 9.0)	1.7
Gunning Fog [48]	6.9 (5.7, 8.4)	2.7
Coleman-Liau Index [26]	5.9 (4.2, 7.4)	3.2
PSK-Dale Chall [107]	5.7 (5.0, 6.5)	1.5
PSK-Flesch [107]	5.4 (5.0, 5.8)	0.8
Flesch Kincaid [66]	5.3 (4.5, 6.6)	2.1
Danielson-Bryan 1 [28]	5.0 (4.7, 5.3)	0.6
New FJP [66]	5.0 (3.0, 6.0)	3.0
PSK-FJP [107]	4.9 (4.6, 5.3)	0.7
PSK-Gunning Fog [107]	4.7 (4.4, 5.1)	0.7
New ARI [66]	4.7 (3.0, 6.2)	3.2
ARI [116]	4.4 (3.1, 5.8)	2.7
Harris Jacobson Wide Range Formula [50]	4.1 (3.4, 6.0)	2.6
Modified SMOG [117]	3.0 (2.0, 5.0)	3.0
Spache Revised [120]	2.4 (2.1, 2.7)	0.6
New Fog Count [66]	1.4 (0.6, 2.7)	2.1

MGL = median grade level; IQR = interquartile range; SMOG = Simple Measure of Gobbledygook; PSK = Powers, Sumner, Kearsley; FJP = Farr, Jenkins, Paterson; ARI = Automated Readability Index.

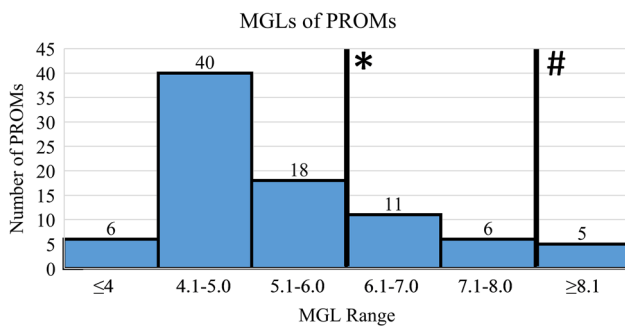


Fig. 1 The median grade level (MGL) distribution of the included independent patient-reported outcome measures (PROMs) are shown. Sixty-four of 86 met the American Medical Association recommendations at or below the 6th grade reading level (black line with *); 81 met the NIH recommendations for the 8th grade reading level (black line with #).

Discussion

PROMs have been increasingly implemented in orthopaedic practice to objectively quantify surgical outcomes and assist in guiding surgical decision making [6, 8, 47]. However, their utility was questioned with a recent report suggesting that most PROMs are written at levels too difficult for the average adult to comprehend [38]. That study is limited by the use of only one readability measure, the Flesch Reading Ease. Our study, using multiple readability measures and giving equal weight to each, seeks to assess the true

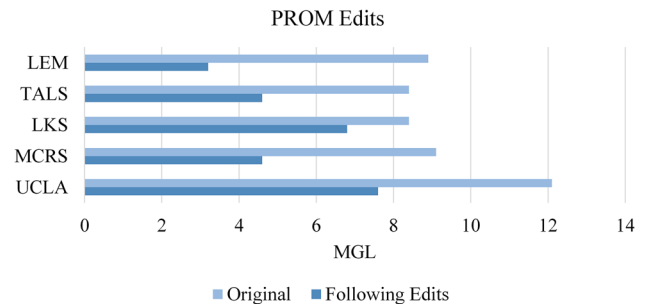


Fig. 2 The median reading grade level (MGL) improvements of low readability PROMs (> 8.0 MGL) after the Centers for Medicare & Medicaid-derived editing process are shown. LEM = Lower Extremity Measure; TALS = Tegner Activity Level Scale; LKS = Lysholm Knee Score; MCRS = Modified Cincinnati Rating System; UCLA = University of California, Los Angeles Activity Score.

readability of orthopaedic-related PROMs. Therefore, we asked: (1) What proportion of orthopaedic-related PROMs and orthopaedic-related portions of the NIH PROMIS® are written at or below the 6th and 8th grade levels? (2) Is there a correlation between the number of questions in the PROM and reading level? (3) Using systematic edits based on guidelines from the CMS [90], what proportion of PROMs achieved NIH-recommended reading levels?

This study has limitations. First, the readability scores were determined by heeding equal weight to each algorithm used. This could be a weakness of the study, as some

formulas could be better equipped and more reliable for use in assessing the readability of healthcare documents, and deserve greater weighting during the determination of MGLs. Additionally, the CMS Toolkit [89] highlights the importance of aesthetics on readability; however, we did not assess such aspects because they could not be analyzed by the software. We also did not evaluate whether the editing process altered the clinical validity and utility of the four selected PROMs. Thus, the possible effects of the editing process on clinical and diagnostic validity merits additional investigation. In addition, this analysis excluded non-English PROMs, as they were unable to be assessed via the readability algorithms used in MGL calculation. Finally, this readability analysis cannot assess the literacy level of PROMs. Readability equations are a numeric method to evaluate PROMs based solely on quantifiable metrics, while literacy involves numerous qualitative factors which this study was not designed to measure. Although having a low MGL does not necessarily translate to higher comprehension and clinical utility, MGLs are the best method currently available to broadly appreciate the level of understanding of healthcare documents among varied patient populations.

The finding that more than 90% of PROMs and all areas of the PROMIS[®] are written at acceptable reading levels refutes the study by El-Daly et al. [38], which led to fears regarding the widespread failure of PROMs. Based on their assessment, only 12% of PROMs had a reading grade level congruent with the average UK literacy level (reported as 11-year-old students or 6th grade), thus questioning the accuracy and reliability of data obtained through PROMs—a sentiment further endorsed in a response by Brown [20]. Inconsistencies between findings in our study and that by El-Daly et al. likely center on their use of a single readability score, the Flesch Reading Ease. While this readability algorithm is mentioned by the CMS, CDC, and NIH as having utility in assessing patient-related documents, it is not, nor is any other readability algorithm, recognized as a gold standard instrument intended to be used in isolation—each entity encourages the use of multiple readability algorithms, not one test in solitude [23, 90, 128]. In our analysis, the Flesch Reading Ease algorithm yielded the third highest MGL of the 19 readability tests used (Table 6). Additionally, the grade level extrapolation of this index score (with original outputs on a scale of 0–100) has only a 5th grade minimum, likely falsely elevating scores, and with an exaggerated baseline [44]. The Flesch Reading Ease is also a dated measure, and questions have arisen regarding its continued utility in assessing health literature [130]. In short, while the Flesch Reading Ease is a commonly used score, its aggressive grade level conversions and lack of adaptation to modern syntax may make it a poor choice on which to base

sweeping PROM readability conclusions, and calls for reform. The potential alarm initiated by El-Daly et al. [38] and endorsed by Brown [20] appears to be overenthusiastic and potentially misleading. However, our findings should be met with guarded optimism. Even though most PROMs are readable to the average American, patients in traditionally low-literacy areas such as the rural southeastern United States where illiteracy rates encompass more than one in three adults [98], may continue to have issues with PROM comprehension. In these areas of decreased literacy, physicians might better serve their patients by selecting PROMs written at a 3rd to 5th grade level [100].

There was no correlation found between numbers of questions in a PROM and associated reading level. While trends of PROM formation are shifting from arduous surveys being multiple pages in length with numerous subsections, to those consisting of short, high-impact questioning [14, 37], it is interesting that reading level is not associated with PROM length. However, the readability algorithms do not assess for possible reader fatigue and document length, but instead analyze sentence and paragraph length. Therefore, while the readability may not be affected in longer, more detailed PROMs, mental fatigue of patients taking the PROMs could play a role. Mental fatigue, studied after traumatic brain injury, has been shown to negatively affect a patient's ability to comprehend new information [62]. Additionally, it has been shown that tired patients are likely to leap to conclusions prematurely [134]. While readability is not affected by PROM length, future research is required to assess the possible effects of reader fatigue on comprehension of the longer PROMs.

Editing according to CMS guidelines improved all PROMs and brought them to or under the 8th grade reading level. These guidelines address many aspects of readability, from text selection to aesthetic appeal; however, in Part 4, Section 3 of the CMS Toolkit, multiple specific suggestions are made, including limiting the number and length of sentences, using the active voice, avoiding acronyms, and using conversational style with nontechnical terms [90]. These were adopted to formulate our editing process (Appendix 2. Supplemental materials are available with the online version of *CORR*[®]) which yielded satisfactory results by lowering the MGL of documents with poor readability by 45%, allowing all to score under the 8th grade reading level (Fig. 2). With the emergence of PROMs as clinical and research tools, steps must be taken to ensure improved readability and sustained validity of measures written over recommended reading levels. Although validation of the CMS-based editing process for use with PROMs is necessary, the improvements in MGL after edits are encouraging. The edited PROMs also would need to be revalidated. Research has shown that minor changes may significantly alter the questions being asked,

and thus, the nature of the responses [94, 105]. While onerous, this revalidation process for these five high-scoring PROMs may be necessary before the edited PROMs are used for clinical research.

PROMs are increasingly used in patient-centered healthcare and outcomes research. Thus, their readability is vital for accurate, valid responses. We disagree with the previous conclusion that the majority of PROMs used in orthopaedics are “incomprehensible to most patients asked to complete them” [38]. In contrast, our study, the most comprehensive analysis of PROM readability to date, revealed that more than 90% of orthopaedic PROMs are written at or below the 8th grade reading level. Additionally, our study tests a method of editing PROMs to reliably decrease the MGL; validation of this method and of edited PROMs is required. Our analysis contradicts previous concerns and provides confidence for the use of nearly all commonly used PROMs in clinical orthopaedic practice.

References

- Agel J, Swiontkowski MF. Guide to outcomes instruments for musculoskeletal trauma research. *J Orthop Trauma*. 2006;20(8 suppl):S1–146.
- Alberta FG, ElAttrache NS, Bissell S, Mohr K, Browdy J, Yocum L, Jobe F. The development and validation of a functional assessment tool for the upper extremity in the overhead athlete. *Am J Sports Med*. 2010;38:903–911.
- Alvey J, Palmer S, Otter S. A comparison of the readability of two patient-reported outcome measures used to evaluate foot surgery. *J Foot Ankle Surg*. 2012;51:412–414.
- Amadio PC, Berquist TH, Smith DK, Ilstrup DM, Cooney WP 3rd, Linscheid RL. Scaphoid malunion. *J Hand Surg Am*. 1989;14:679–687.
- American Academy of Orthopaedic Surgeons. Patient Reported Outcome Measures. Available at: http://www.aaos.org/quality/performance_measures/patient_reported_outcome_measures/. Accessed November 20, 2016.
- Ayers DC. Implementation of patient-reported outcome measures in total knee arthroplasty. *J Am Acad Orthop Surg*. 2017;25(suppl 1):S48–50.
- Badarudeen S, Sabharwal S. Assessing readability of patient education materials: current role in orthopaedics. *Clin Orthop Relat Res*. 2010;468:2572–2580.
- Baumhauer JF, Bozic KJ. Value-based healthcare: patient-reported outcomes in clinical decision making. *Clin Orthop Relat Res*. 2016;474:1375–1378.
- Beaton DE, Wright JG, Katz JN. Development of the Quick-DASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am*. 2005;87:1038–1046.
- Behrend H, Giesinger K, Giesinger JM, Kuster MS. The “forgotten joint” as the ultimate goal in joint arthroplasty: validation of a new patient-reported outcome measure. *J Arthroplasty*. 2012;27:430–436.e431.
- Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol*. 1988;15:1833–1840.
- Bennett PJ, Patterson C, Wearing S, Baglioni T. Development and validation of a questionnaire designed to measure foot-health status. *J Am Podiatr Med Assoc*. 1998;88:419–428.
- Binkley JM, Stratford PW, Lott SA, Riddle DL. 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.
- Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 2013;346:f167.
- Bolton JE, Breen AC. The Bournemouth Questionnaire: a short-form comprehensive outcome measure. I. Psychometric properties in back pain patients. *J Manipulative Physiol Ther*. 1999;22:503–510.
- Bolton JE, Humphreys BK. The Bournemouth Questionnaire: a short-form comprehensive outcome measure. II. Psychometric properties in neck pain patients. *J Manipulative Physiol Ther*. 2002;25:141–148.
- Bormuth JR. Readability: a new approach. *Reading Res Q*. 1966;1:79–132.
- Bourne RB. Measuring tools for functional outcomes in total knee arthroplasty. *Clin Orthop Relat Res*. 2008;466:2634–2638.
- Bremander AB, Petersson IF, Roos EM. Validation of the Rheumatoid and Arthritis Outcome Score (RAOS) for the lower extremity. *Health Qual Life Outcomes*. 2003;1:55.
- Brown TD. CORR Insights®: Are patient-reported outcome measures in orthopaedics easily read by patients? *Clin Orthop Relat Res*. 2016;474:256–257.
- Budiman-Mak E, Conrad K, Stuck R, Matters M. Theoretical model and Rasch analysis to develop a revised Foot Function Index. *Foot Ankle Int*. 2006;27:519–527.
- Caylor J, Sticht T, Fox L, Ford J. Methodologies for determining reading requirements of military occupational specialties. Human Resources Research Organization. Available at: <http://files.eric.ed.gov/fulltext/ED074343.pdf>. Accessed December 9, 2016.
- Centers for Disease Control and Prevention, U.S. Department of Health and Human Services. Simply Put: A guide for creating easy-to-understand materials. Available at: http://www.cdc.gov/HealthLiteracy/pdf/Simply_Put.pdf. Accessed November 25, 2016.
- Choudhry AJ, Baghdadi YM, Wagie AE, Habermann EB, Heller SF, Jenkins DH, Cullinane DC, Zielinski MD. Readability of discharge summaries: with what level of information are we dismissing our patients? *Am J Surg*. 2016;211:631–636.
- Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. *J Hand Surg Am*. 1998;23:575–587.
- Coleman M, Liau TL. A computer readability formula designed for machine scoring. *J Appl Psychol*. 1975;60:283.
- Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. *Clin Orthop Relat Res*. 1987;214:160–164.
- Danielson WA, Bryan SD. Computer automation of two readability formulas. *Journalism Mass Commun Q*. 1963;40:201–206.
- Davis AM, Perruccio AV, Canizares M, Tennant A, Hawker GA, Conaghan PG, Roos EM, Jordan JM, Maillefert JF, Dougados M, Lohmander LS. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. *Osteoarthritis Cartilage*. 2008;16:551–559.
- Dawson J, Doll H, Boller I, Fitzpatrick R, Little C, Rees J, Jenkinson C, Carr AJ. The development and validation of a patient-reported questionnaire to assess outcomes of elbow surgery. *J Bone Joint Surg Br*. 2008;90:466–473.

31. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg Br.* 1996;78:593–600.
32. Dawson J, Fitzpatrick R, Carr A. The assessment of shoulder instability: the development and validation of a questionnaire. *J Bone Joint Surg Br.* 1999;81:420–426.
33. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br.* 1996;78:185–190.
34. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg Br.* 1998;80:63–69.
35. De Oliveira GS Jr, McCarthy RJ, Wolf MS, Holl J. The impact of health literacy in the care of surgical patients: a qualitative systematic review. *BMC Surg.* 2015;15:86.
36. Domsic RT, Saltzman CL. Ankle osteoarthritis scale. *Foot Ankle Int.* 1998;19:466–471.
37. Edwards P, Roberts I, Clarke M, DiGiuseppi C, Pratap S, Wentz R, Kwan I. Increasing response rates to postal questionnaires: systematic review. *BMJ.* 2002;324:1183.
38. El-Daly I, Ibraheim H, Rajakulendran K, Culpan P, Bates P. Are patient-reported outcome measures in orthopaedics easily read by patients? *Clin Orthop Relat Res.* 2016;474:246–255.
39. EuroQol Group. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy.* 1990;16:199–208.
40. EuroQol Research Foundation. How to obtain EQ-5D. Available at: <http://www.euroqol.org/eq-5d-products/how-to-obtain-eq-5d.html>. Accessed December 7, 2016.
41. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy.* 1980;66:271–273.
42. Fairbank JC, Pynsent PB. The Oswestry Disability Index. *Spine (Phila Pa 1976).* 2000;25:2940–2952; discussion 2952.
43. Flesch RF. A new readability yardstick. *J Appl Psychol.* 1948;32:221–233.
44. Flesch RF. *How to Write Plain English: A Book for Lawyers and Consumers.* New York, NY: Harper and Row; 1979.
45. Friedman DB, Hoffman-Goetz L. A systematic review of readability and comprehension instruments used for print and web-based cancer information. *Health Educ Behav.* 2006;33:352–373.
46. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum.* 1980;23:137–145.
47. Greene ME, Rolfson O, Gordon M, Garellick G, Nemes S. Standard comorbidity measures do not predict patient-reported outcomes 1 year after total hip arthroplasty. *Clin Orthop Relat Res.* 2015;473:3370–3379.
48. Gunning R. *The Technique of Clear Writing.* New York, NY: McGraw-Hill; 1952.
49. Hale SA, Hertel J. Reliability and sensitivity of the Foot and Ankle Disability Index in subjects with chronic ankle instability. *J Athl Train.* 2005;40:35–40.
50. Harris AJ, Jacobson MD. *Basic Reading Vocabularies.* New York, NY: Macmillan; 1982.
51. Hildebrand KA, Buckley RE, Mohtadi NG, Faris P. Functional outcome measures after displaced intra-articular calcaneal fractures. *J Bone Joint Surg Br.* 1996;78:119–123.
52. Horner SD, Surratt D, Juliusson S. Improving readability of patient education materials. *J Community Health Nurs.* 2000;17:15–23.
53. 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.
54. Hudson-Cook N, Tomes-Nicholson K, Breen A. Revised Oswestry disability questionnaire. In: Roland M, Jenner J, eds. *Back Pain: New Approaches to Rehabilitation and Education.* New York, NY: Manchester University Press; 1989:187–204.
55. Hunt SM, McKenna SP, McEwen J, Backett EM, Williams J, Papp E. A quantitative approach to perceived health status: a validation study. *J Epidemiol Community Health.* 1980;34:281–286.
56. Insall JN, Dorr LD, Scott RD, Scott WN. Rationale of the Knee Society clinical rating system. *Clin Orthop Relat Res.* 1989;248:13–14.
57. Irrgang JJ, Anderson AF, Boland AL, Harner CD, Kurosaka M, Neyret P, Richmond JC, Shelborne KD. Development and validation of the international knee documentation committee subjective knee form. *Am J Sports Med.* 2001;29:600–613.
58. Irrgang JJ, Snyder-Mackler L, Wainner RS, Fu FH, Harner CD. Development of a patient-reported measure of function of the knee. *J Bone Joint Surg Am.* 1998;80:1132–1145.
59. Jaglal S, Lakhani Z, Schatzker J. Reliability, validity, and responsiveness of the lower extremity measure for patients with a hip fracture. *J Bone Joint Surg Am.* 2000;82:955–962.
60. Johanson NA, Charlson ME, Szatrowski TP, Ranawat CS. A self-administered hip-rating questionnaire for the assessment of outcome after total hip replacement. *J Bone Joint Surg Am.* 1992;74:587–597.
61. Johanson NA, Liang MH, Daltroy L, Rudicel S, Richmond J. American Academy of Orthopaedic Surgeons lower limb outcomes assessment instruments: reliability, validity, and sensitivity to change. *J Bone Joint Surg Am.* 2004;86:902–909.
62. Johansson B, Berglund P, Ronnback L. Mental fatigue and impaired information processing after mild and moderate traumatic brain injury. *Brain Inj.* 2009;23:1027–1040.
63. Jordan A, Manniche C, Mosdal C, Hindsberger C. The Copenhagen Neck Functional Disability Scale: a study of reliability and validity. *J Manipulative Physiol Ther.* 1998;21:520–527.
64. Juul T, Sogaard K, Roos EM, Davis AM. Development of a patient-reported outcome: the Neck OutCome Score (NOOS): content and construct validity. *J Rehabil Med.* 2015;47:844–853.
65. Kaikkonen A, Kannus P, Jarvinen M. A performance test protocol and scoring scale for the evaluation of ankle injuries. *Am J Sports Med.* 1994;22:462–469.
66. Kincaid JP, Fishburne RP Jr, Rogers RL, Chissom BS. Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel. Naval Air Station Memphis, Millington, Tennessee. Available at: <http://www.dtic.mil/dtic/tr/fulltext/u2/a006655.pdf>. Accessed December 9, 2016.
67. King GJ, Richards RR, Zuckerman JD, Blasler R, Dillman C, Friedman RJ, Gartsman GM, Iannotti JP, Murnahan JP, Mow VC, Woo SL. A standardized method for assessment of elbow function. Research Committee, American Shoulder and Elbow Surgeons. *J Shoulder Elbow Surg.* 1999;8:351–354.
68. Kirkley A, Alvarez C, Griffin S. The development and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: the Western Ontario Rotator Cuff Index. *Clin J Sport Med.* 2003;13:84–92.
69. Kirkley A, Griffin S, McLintock H, Ng L. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability: the Western Ontario Shoulder Instability Index (WOSI). *Am J Sports Med.* 1998;26:764–772.
70. Kirkley A, Griffin S, Whelan D. The development and validation of a quality of life-measurement tool for patients with meniscal pathology: the Western Ontario Meniscal Evaluation Tool (WOMET). *Clin J Sport Med.* 2007;17:349–356.
71. Kohn D, Geyer M. The subjective shoulder rating system. *Arch Orthop Trauma Surg.* 1997;116:324–328.

72. Kopec JA, Esdaile JM, Abrahamowicz M, Abenheim L, Wood-Dauphinee S, Lamping DL, Williams JI. The Quebec Back Pain Disability Scale: measurement properties. *Spine (Phila Pa 1976)*. 1995;20:341–352.
73. Kujala UM, Jaakkola LH, Koskinen SK, Taimela S, Hurme M, Nelimarkka O. Scoring of patellofemoral disorders. *Arthroscopy*. 1993;9:159–163.
74. Kurer M, Gooding C. Orthopaedic Scores. Available at: <http://www.orthopaedicscores.com/>. Accessed November 20, 2016.
75. L'Insalata JC, Warren RF, Cohen SB, Altchek DW, Peterson MG. A self-administered questionnaire for assessment of symptoms and function of the shoulder. *J Bone Joint Surg Am*. 1997;79:738–748.
76. Lawlis GF, Cuencas R, Selby D, McCoy CE. The development of the Dallas Pain Questionnaire: an assessment of the impact of spinal pain on behavior. *Spine (Phila Pa 1976)*. 1989;14:511–516.
77. Lequesne MG, Mery C, Samson M, Gerard P. Indexes of severity for osteoarthritis of the hip and knee: validation: value in comparison with other assessment tests. *Scand J Rheumatol Suppl*. 1987;65:85–89.
78. Levine DW, Simmons BP, Koris MJ, Daltroy LH, Hohl GG, Fossel AH, Katz JN. 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.
79. Lo IK, Griffin S, Kirkley A. The development of a disease-specific quality of life measurement tool for osteoarthritis of the shoulder: the Western Ontario Osteoarthritis of the Shoulder (WOOS) index. *Osteoarthritis Cartilage*. 2001;9:771–778.
80. Lyman S, Lee YY, Franklin PD, Li W, Cross MB, Padgett DE. Validation of the KOOS, JR: a short-form knee arthroplasty outcomes survey. *Clin Orthop Relat Res*. 2016;474:1461–1471.
81. Lyman S, Lee YY, Franklin PD, Li W, Mayman DJ, Padgett DE. Validation of the HOOS, JR: a short-form hip replacement survey. *Clin Orthop Relat Res*. 2016;474:1472–1482.
82. Lysholm J, Gillquist J. Evaluation of knee ligament surgery results with special emphasis on use of a scoring scale. *Am J Sports Med*. 1982;10:150–154.
83. MacDermid JC, Turgeon T, Richards RS, Beadle M, Roth JH. Patient rating of wrist pain and disability: a reliable and valid measurement tool. *J Orthop Trauma*. 1998;12:577–586.
84. Majeed SA. Grading the outcome of pelvic fractures. *J Bone Joint Surg Br*. 1989;71:304–306.
85. Martin RL, Irrgang JJ, Burdett RG, Conti SF, Van Swearingen JM. Evidence of validity for the Foot and Ankle Ability Measure (FAAM). *Foot Ankle Int*. 2005;26:968–983.
86. Marx RG, Stump TJ, Jones EC, Wickiewicz TL, Warren RF. Development and evaluation of an activity rating scale for disorders of the knee. *Am J Sports Med*. 2001;29:213–218.
87. Matsen FA 3rd, Ziegler DW, DeBartolo SE. Patient self-assessment of health status and function in glenohumeral degenerative joint disease. *J Shoulder Elbow Surg*. 1995;4:345–351.
88. Matsumoto M, Baba T, Homma Y, Kobayashi H, Ochi H, Yuasa T, Behrend H, Kaneko K. Validation study of the Forgotten Joint Score-12 as a universal patient-reported outcome measure. *Eur J Orthop Surg Traumatol*. 2015;25:1141–1145.
89. McGee J, Centers for Medicare and Medicaid Services, U.S. Department of Health and Human Services. Toolkit for Making Written Material Clear and Effective, Section 4: Special Topics for Writing and Design. Available at: <https://www.cms.gov/Outreach-and-Education/Outreach/WrittenMaterialsToolkit/downloads/ToolkitPart07.pdf>. Accessed November 25, 2016.
90. McGee J, Centers for Medicare and Medicaid Services, U.S. Department of Health and Human Services. Toolkit Part 4: Guidelines for Writing. Available at: <https://www.cms.gov/Outreach-and-Education/Outreach/WrittenMaterialsToolkit/ToolkitPart04.html>. Accessed November 20, 2016.
91. McLaughlin HG. SMOG grading: a new readability formula. *J Reading*. 1969;12:639–646.
92. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. *Pain*. 1975;1:277–299.
93. Melzack R. The short-form McGill Pain Questionnaire. *Pain*. 1987;30:191–197.
94. Miller PR. Tipsheet: Question Wording. Duke Initiative on Survey Methodology. Available at: <http://www.dism.srri.duke.edu/pdfs/Tipsheet%20-%20Question%20Wording.pdf>. Accessed February 24, 2017.
95. Morgan S. Miscommunication between patients and general practitioners: implications for clinical practice. *J Prim Health Care*. 2013;5:123–128.
96. Morrey B. Functional Evaluation of the Elbow. In Morrey B, ed. *The Elbow and its Disorders*. Philadelphia, PA: WB Saunders Co; 2000:74–83.
97. Naal FD, Hatzung G, Muller A, Impellizzeri F, Leunig M. Validation of a self-reported Beighton score to assess hypermobility in patients with femoroacetabular impingement. *Int Orthop*. 2014;38:2245–2250.
98. National Center for Education Statistics, U.S. Department of Education. National Assessment of Adult Literacy: State and County Estimates of Low Literacy. Available at: <https://nces.ed.gov/naal/estimates/StateEstimates.aspx>. Accessed February 24, 2017.
99. National Center for Health Marketing, Centers for Disease Control and Prevention. What We Know About Health Literacy. Available at: <http://www.cdc.gov/healthcommunication/pdf/healthliteracy.pdf>. Accessed November 20, 2016.
100. National Institutes of Health, U.S. Department of Health and Human Services. Clear & Simple: What is Clear & Simple?. Available at: <https://www.nih.gov/institutes-nih/nih-office-director/office-communications-public-liaison/clear-communication/clear-simple>. Accessed February 24, 2017.
101. Nilsson AK, Lohmander LS, Klassbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS): validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord*. 2003;4:10.
102. Noyes FR, Barber SD, Mooar LA. A rationale for assessing sports activity levels and limitations in knee disorders. *Clin Orthop Relat Res*. 1989;246:238–249.
103. Pace CC, Atcherson SR, Zraick RI. A computer-based readability analysis of patient-reported outcome questionnaires related to oral health quality of life. *Patient Educ Couns*. 2012;89:76–81.
104. Perruccio AV, Stefan Lohmander L, Canizares M, Tennant A, Hawker GA, Conaghan PG, Roos EM, Jordan JM, Maillefert JF, Dougados M, Davis AM. The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS): an OARSI/OMERACT initiative. *Osteoarthritis Cartilage*. 2008;16:542–550.
105. Pew Research Center. Questionnaire Design. Available at: <http://www.pewresearch.org/methodology/u-s-survey-research/questionnaire-design/>. Accessed February 24, 2017.
106. Pincus T, Summey JA, Soraci SA Jr, Wallston KA, Hummon NP. Assessment of patient satisfaction in activities of daily living using a modified Stanford Health Assessment Questionnaire. *Arthritis Rheum*. 1983;26:1346–1353.
107. Powers RD, Sumner WA, Kearn BE. A recalculation of four adult readability formulas. *J Educ Psychol*. 1958;49:99.
108. Ramkumar PN, Harris JD, Noble PC. Patient-reported outcome measures after total knee arthroplasty: a systematic review. *Bone Joint Res*. 2015;4:120–127.

109. Richards RR, An KN, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, Iannotti JP, Mow VC, Sidles JA, Zuckerman JD. A standardized method for the assessment of shoulder function. *J Shoulder Elbow Surg.* 1994;3:347–352.
110. Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. *Arthritis Care Res.* 1991;4:143–149.
111. Roland M, Morris R. A study of the natural history of back pain: Part I. development of a reliable and sensitive measure of disability in low-back pain. *Spine (Phila Pa 1976).* 1983;8:141–144.
112. Roos EM, Brandsson S, Karlsson J. Validation of the foot and ankle outcome score for ankle ligament reconstruction. *Foot Ankle Int.* 2001;22:788–794.
113. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS): development of a self-administered outcome measure. *J Orthop Sports Phys Ther.* 1998;28:88–96.
114. Saleh KJ, Mulhall KJ, Bershadsky B, Ghomrawi HM, White LE, Buyea CM, Krackow KA. Development and validation of a lower-extremity activity scale: use for patients treated with revision total knee arthroplasty. *J Bone Joint Surg Am.* 2005;87:1985–1994.
115. Sheppard ED, Hyde Z, Florence MN, McGwin G, Kirchner JS, Ponce BA. Improving the readability of online foot and ankle patient education materials. *Foot Ankle Int.* 2014;35:1282–1286.
116. Smith EA, Senter R. Automated readability index. Aerospace Medical Research Laboratories, Aerospace Medical Division, Air Force Systems Command. Available at: <http://www.dtic.mil/dtic/tr/fulltext/u2/667273.pdf>. Accessed December 9, 2016.
117. Smith LL. Using a modified SMOG in primary and intermediate grades. *Reading Horizons.* 1984;24:129–132.
118. Smith SG, Curtis LM, Wardle J, von Wagner C, Wolf MS. Skill set or mind set? Associations between health literacy, patient activation and health. *PLoS One.* 2013;8:e74373.
119. Smith SK, Dixon A, Trevena L, Nutbeam D, McCaffery KJ. Exploring patient involvement in healthcare decision making across different education and functional health literacy groups. *Soc Sci Med.* 2009;69:1805–1812.
120. Spache G. A new readability formula for primary-grade reading materials. *The Elementary School Journal.* 1953;53:410–413.
121. Stewart AL, Hays RD, Ware JE Jr. The MOS short-form general health survey. Reliability and validity in a patient population. *Med Care.* 1988;26:724–735.
122. Swiontkowski MF, Engelberg R, Martin DP, Agel J. Short musculoskeletal function assessment questionnaire: validity, reliability, and responsiveness. *J Bone Joint Surg Am.* 1999;81:1245–1260.
123. Tartaglione JP, Rosenbaum AJ, Abousayed M, Hushmendi SF, DiPrea JA. Evaluating the quality, accuracy, and readability of online resources pertaining to hallux valgus. *Foot Ankle Spec.* 2016;9:17–23.
124. Tegner Y, Lysholm J. Rating systems in the evaluation of knee ligament injuries. *Clin Orthop Relat Res.* 1985;198:43–49.
125. Templeman D, Goulet J, Duwelius PJ, Olson S, Davidson M. Internal fixation of displaced fractures of the sacrum. *Clin Orthop Relat Res.* 1996;329:180–185.
126. Thorborg K, Holmich P, Christensen R, Petersen J, Roos EM. The Copenhagen Hip and Groin Outcome Score (HAGOS): development and validation according to the COSMIN checklist. *Br J Sports Med.* 2011;45:478–491.
127. U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. National Action Plan to Improve Health Literacy. Available at: https://health.gov/communication/hlactionplan/pdf/Health_Literacy_Action_Plan.pdf. Accessed November 20, 2016.
128. U.S. National Library of Medicine, National Institutes of Health, U.S. Department of Health and Human Services. How to Write Easy-to-Read Health Materials. Available at: <https://medlineplus.gov/etr.html>. Accessed November 20, 2016.
129. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. *J Manipulative Physiol Ther.* 1991;14:409–415.
130. Wang LW, Miller MJ, Schmitt MR, Wen FK. Assessing readability formula differences with written health information materials: application, results, and recommendations. *Res Social Adm Pharm.* 2013;9:503–516.
131. Ware JE Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care.* 1996;34:220–233.
132. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. conceptual framework and item selection. *Med Care.* 1992;30:473–483.
133. Washburn RA, Smith KW, Jette AM, Janney CA. The Physical Activity Scale for the Elderly (PASE): development and evaluation. *J Clin Epidemiol.* 1993;46:153–162.
134. Webster DM, Richter L, Kruglanski AW. On leaping to conclusions when feeling tired: mental fatigue effects on impression primacy. *J Exper Social Psychol.* 1996;32:181–195.
135. Weiss BD. Health Literacy and Patient Safety: Help Patients Understand. Manual for Clinicians. 2nd ed. American Medical Association Foundation and the American Medical Association. Available at: http://med.fsu.edu/userFiles/file/ahec_health_clinicians_manual.pdf. Accessed November 20, 2016.
136. Zahiri CA, Schmalzried TP, Szuszczewicz ES, Amstutz HC. Assessing activity in joint replacement patients. *J Arthroplasty.* 1998;13:890–895.