

Original Publication

OPEN ACCESS

Assessing Radiographic Knee Osteoarthritis: An Online Training Tutorial for the Kellgren-Lawrence Grading Scale

Bethany Hayes*, Andrew Kittelson, PT, DPT, Brian Loyd, PT, DPT, Elizabeth Wellsandt, PT, DPT, OCS, Jonathan Flug, MBA, MD, Jennifer Stevens-Lapsley, PT, PhD

*Corresponding author: hayes.bethanyd@gmail.com

Citation: Hayes B, Kittelson A, Loyd B, Wellsandt E, Flug J, Stevens-Lapsley J. Assessing radiographic knee osteoarthritis: an online training tutorial for the Kellgren-Lawrence grading scale. *MedEdPORTAL*. 2016;12:10503. https://doi.org/10.15766/mep_2374-8265.10503

Copyright: © 2016 Hayes et al. This is an open-access publication distributed under the terms of the Creative Commons Attribution-NonCommercial-Share Alike license.

Abstract

Introduction: Knee osteoarthritis (OA) is a degenerative form of arthritis commonly diagnosed in older adults. It presents clinically with patient complaints of pain and impaired function, which are thought to result from cartilage degeneration and other skeletal changes. These changes can be examined radiographically and quantified using the semiquantitative grading scale known as the Kellgren-Lawrence (KL) scale. Currently, no standard training exists for KL grading, which may explain the unsatisfactory reliability of this tool in OA research. Therefore, the objective of this project was to develop a training tutorial for KL grading of knee OA to educate assessors on possible areas of inconsistency in grading. **Methods:** The tutorial was developed in an e-learning authoring tool, Articulate Presenter. The content focuses on the poor reliability of KL grading, normal anatomy of a knee radiograph, and multiple examples of bony changes within the knee and their relation to different grades of the KL scale. The tutorial was presented to a group of health sciences graduate students at the University of Colorado Denver. **Results:** Students were able to complete the training and an associated assessment in under an hour and reported improved confidence with assessing radiographic knee OA. Furthermore, they demonstrated favorable inter- and intrarater reliability scores in applying KL grading. **Discussion:** To our knowledge, this is the first attempt to standardize training in KL grading for knee OA and to examine the effects of this training on reliability.

Keywords

Editor's Choice, Osteoarthritis, Physical and Rehabilitation Medicine, Kellgren-Lawrence, Knee Osteoarthritis, Knee Reliability, Weighted Kappa Statistics

Educational Objectives

By the end of this session, learners will be able to:

1. Describe the features of joint degeneration commonly observed in knee osteoarthritis.
2. Describe the poor reliability associated with the Kellgren-Lawrence grading scale.
3. Compare and contrast the descriptions of the Kellgren-Lawrence grading scale.
4. Identify signs of osteoarthritis when presented with a knee radiograph.
5. Apply an accurate Kellgren-Lawrence grade when presented with a knee radiograph.

Introduction

Knee osteoarthritis (OA) is one of the most common musculoskeletal disorders seen in older adults. It is often associated with maladaptive joint loading, either from repetitive high-impact activities or because of increased load due to obesity.¹⁻⁴ Skeletal changes, thought to occur in response to this physical stress and other factors, are most often assessed by radiography. These radiographic changes (primarily the formation of osteophytes and thinning of articular cartilage) have been codified in the 5-point, semiquantitative Kellgren-Lawrence (KL) grading scale, the use of which is nearly ubiquitous in OA clinical research.¹⁻⁶ However, despite being the most widely used tool for grading OA severity since its development in 1957,⁷ the KL scale has historically demonstrated inconsistent and unsatisfactory

Appendices

- A. Kellgren-Lawrence Grading Scale Tutorial folder
- B. Kellgren-Lawrence Grading Scale Self-Assessment.pdf
- C. Kellgren-Lawrence Grading Scale Answer Key.pdf

All appendices are peer reviewed as integral parts of the Original Publication.

reliability.^{2,3,6} Even among experienced graders, interrater agreement (as measured by a weighted kappa statistic) has been seen to be as low as .36, which is regarded as poor reliability.⁶

The reasons underlying the poor reliability of the KL scale are rarely discussed and are often attributed to the subjectivity of the scale itself.⁴ Another distinct—but unexplored—possibility is that graders lack sufficient training to apply the scale consistently. Clinical experience alone is often considered adequate preparation for applying KL grades, and other training methods involve informal discussions among the one or two graders involved in a research project.^{2,4,6} There are likely to be inconsistent interpretations of the KL scale across graders and studies.³ The purpose of this work was to develop a training tutorial for KL grading of knee OA, to refine the tutorial in response to feedback from users regarding the tutorial's usefulness and interpretability, and to test the inter- and intrarater reliabilities of KL grading among students trained by the tutorial, in order to compare them to reliability reported in the knee OA literature.

Methods

The original grade definitions of the KL scale are used in this educational tutorial: Grade 0: no pathological features; Grade 1: doubtful narrowing of joint space and possible osteophytic lipping; Grade 2: definite osteophytes and possible narrowing of joint space; Grade 3: moderate multiple osteophytes, definite narrowing of joint space, some sclerosis, and possible deformity of bony ends; Grade 4: large osteophytes, marked narrowing of joint space, severe sclerosis, and definite deformity of bone ends.⁷ Anterior-posterior weight-bearing radiographs were used for the tutorial and quiz.

A training tutorial titled *The Kellgren-Lawrence Grading Scale: An Online Training Tutorial for Knee Radiographs* (Appendix A) was developed in an e-learning authoring tool (Articulate Presenter) through a four-phase iterative process involving feedback from multiple possible end-user groups: clinical researchers, radiologists, and graduate students. The tutorial was initially crafted in consultation with an experienced user of KL grades for clinical knee OA research and focused on three major areas: background on knee OA, anatomy, and the KL scale; detailed descriptions of KL Grades 0-4, with examples; and extensive discussion of possible sources of grading inconsistencies, with multiple illustrative examples. Special attention was paid to the distinction between possible and definite osteophytes, as well as possible versus definite joint space narrowing, as these features are key for distinguishing KL grades.

The training tutorial was then iteratively refined with feedback from three possible user groups: clinical researchers, radiologists, and health sciences graduate students. The initial draft was viewed by a group of five clinical researchers. Their feedback included a suggestion to add the definition of each grade to its associated slide to reinforce the relevant aspects of each definition and to condense the background to give more time to explore and discuss the radiographic images. The background was condensed to provide only essential information needed to understand knee OA in relation to the KL grading scale. The second draft was viewed by four graduate students. Feedback targeted the need to incorporate a mini-quiz for self-testing as well as to add a clearer distinction of possible versus definite osteophytes. These suggestions led to the adoption of a strategy of encouraging student reflection during the tutorial by displaying a radiograph, asking the student to identify key features and determine the correct grade, and then pausing before providing the results. An additional slide was designed to provide detailed information about the difference between possible and definite osteophytes since this distinction is key in KL grading. Following these modifications, the third draft was reviewed by an experienced radiologist, who suggested deemphasizing radiographic features (e.g., sclerosis) not directly pertinent to KL grading. The feedback from the radiologist was incorporated, and sclerosis was only briefly mentioned for definitional purposes; this product was used as the final tutorial for testing reliability.

The training tutorial was tested for usability and preliminary effectiveness in a group of 47 health sciences graduate students at the University of Colorado Anschutz Medical Campus. All 47 students had no prior experience with the KL grading scale but had been exposed to basic radiography. After participating in

the training tutorial, all participants completed an assessment, wherein they were asked to apply KL grades to 30 unique knee radiographs, so that interrater reliability statistics could be calculated. Fifteen duplicated radiographs were also randomized throughout the assessment. The duplicated radiographs were used for calculation of intrarater reliability statistics. Inter- and intrarater reliabilities of grading of radiographic knee OA were calculated using weighted kappa statistics (κ_w). Inter- and intrarater reliabilities were compared between groups using analysis of variance (ANOVA) to determine if the experimental training tutorial resulted in improved reliability relative to industry standards. Weighted kappa statistics using quadratic weighting were calculated and averaged for all unique combinations of raters (interrater reliability) or each rater, Time 1 versus Time 2 (intrarater reliability). All groups were put into a one-way ANOVA with a kappa statistic as the outcome. Values were calculated using the Statistical Analysis Software (Version 9.4).

To access the tutorial in Appendix A, download the content onto a computer, and open the folder. The folder has several files. Open the file named presentation.html, and a web browser window will open the module. The tutorial will begin automatically, but the slide will not advance until Next is clicked. Each slide can be replayed by dragging the slider at the bottom of the screen to the beginning of the time bar or pressing the replay button. To go to a previous slide, click the Previous button. The KL grading scale file is meant to be used in conjunction with the tutorial; a note-taking section under the grading scale allows users to make notes while working through the slides. After the tutorial is completed, a quiz (Appendix B) is available for practice. The quiz features one radiograph per page for learners to grade. An answer key (Appendix C) with explanations of each radiograph allows learners to check their answers.

Results

The content of the tutorial is the result of learning and defining the KL grading scale and refining presentation of the material over three rounds of feedback from possible user groups. A common critique was that the distinction between KL Grades 3 and 4 was not stressed sufficiently, but ultimately, student feedback on the final version was positive and focused on several themes:

1. Students commented positively on the reiteration of the major ideas and the pictorial comparisons of each KL grade.
 - “I think having the examples was very helpful, and the comparison between the grades was enlightening. I felt like I knew how to use the Kellgren-Lawrence rating scale.”
 - “I found it helpful to have the important points repeated multiple times. I also liked seeing examples of the different grades and listening to the reasoning about how to determine the grade. It was good to see the grades compared to the grades above and below to help differentiate between the two.”
 - “I appreciated the repetition of ‘first, look at the osteophytes, then the narrowing of space. . . .’ I felt like I understood the method and order well.”
2. Students provided positive feedback regarding the questions asked during the tutorial.
 - “I liked the practice for grading that I was able to get throughout the tutorial. There was always a reinforcement of what I learned. I liked that it wasn’t as overwhelming as I thought it would be. It made me feel as though grading knees wasn’t too hard, but I definitely understood that the KL scale was subjective.”
 - “The practice questions were in sufficient quantity that I got adequate practice identifying pathologies.”
 - “I also loved how there were interactive questions throughout to keep the audience’s attention.”

3. Students' confidence in KL grading improved following the training.
 - “This learning tutorial really helped me with my understanding of knee osteoarthritis. I knew nothing about the Kellgren-Lawrence grading scale before and now I'm sure I can assess a knee with osteoarthritis pretty confidently. I love how the tutorial had specific key points that someone can use to rate the grading scale such as the characteristic of an osteophyte or the size of joint space. Those factors really helped me as I was taking the quiz.”
4. Overall feedback was positive.
 - “The tutorial was concise and informative. I like that. I feel I learned a lot in a short period of time.”
 - “Excellent, clean and clear slides. Very well paced narration and explanations. Extremely informative, and key points are reiterated multiple times to reinforce learning. Fantastic job!”
 - “I also liked this tutorial because I really felt like I learned something new (and it wasn't too overwhelming to get there) and it was very interesting! To me, it was worth the time to take the tutorial!”

These comments highlight the helpfulness of the tutorial and ultimately the confidence in abilities after watching the tutorial. One student said, “I can see how without training, this would be confusing.” Plans are in development to implement the tutorial in the radiology residency program at the University of Colorado Anschutz Medical Campus.

Weighted kappa values of 0 indicate poor agreement, values of .01-.20 indicate slight agreement, values of .21-.40 indicate fair agreement, values of .41-.60 indicate moderate agreement, values of .61-.80 indicate substantial agreement, and values of .81-.99 indicate almost perfect agreement.⁸ Interrater reliability of KL grading among the final group of students ($k_w = .82$) and intrarater reliability of KL grading in the experimental group ($k_w = .85$) were almost perfect and compare favorably with weighted kappa values for experienced graders that have been reported in the literature.⁶ Therefore, at the very least, this study suggests that utilization of a training tutorial for the KL grading scale results in interrater reliability that is favorable, compared to what has been observed in knee OA research.⁶

Discussion

To our knowledge, this is the first attempt to standardize training in KL grading for knee OA and to examine the effects of this training on reliability. Interrater reliability was excellent among participants who received the training tutorial, although further testing with a control group or pre-post assessments should be conducted to determine if reliability can be attributed directly to the training tutorial.

The tutorial helps promote a standardized approach to KL grading. In recognition of the subjectivity of the KL scale, the tutorial more clearly defines the grades and associated features to improve reliability of grading. One limitation of this study is that weight-bearing radiographs are used; however, since weight-bearing radiographs are the standard approach in clinical practice, this proves to be a minor issue. Future iterations of the training tutorial could be developed to improve reliability by including a better delineation between KL Grades 3 and 4, as several students reported difficulty with this particular aspect of KL grading. This improvement would benefit learners by stressing the differences between these two grades and hopefully further increasing reliability.

Bethany Hayes: Graduate of the Modern Human Anatomy Master's Program, University of Colorado School of Medicine Anschutz Medical Campus

Andrew Kittelson, PT, DPT: Graduate of the Rehabilitation Sciences PhD Program, University of Colorado School of Medicine Anschutz Medical Campus

Brian Loyd, PT, DPT: Student in the Rehabilitation Sciences PhD Program, University of Colorado School of Medicine Anschutz Medical Campus

Elizabeth Wellsandt, PT, DPT, OCS: Student in the Biomechanics and Movement Science PhD Program, University of Delaware

Jonathan Flug, MBA, MD: Assistant Professor, Department of Musculoskeletal Radiology, University of Colorado School of Medicine Anschutz Medical Campus

Jennifer Stevens-Lapsley, PT, PhD: Professor, Physical Therapy Department, University of Colorado School of Medicine Anschutz Medical Campus

Disclosures

None to report.

Funding/Support

None to report.

Ethical Approval

This publication contains data obtained from human subjects and received ethical approval.

References

1. Paradowski PT. Osteoarthritis of the knee: assessing the disease. *Health Care Curr Rev.* 2014;2(2):e103.
 2. Culvenor AG, Engen CN, Oiestad BE, Engebretsen L, Risberg MA. Defining the presence of radiographic knee osteoarthritis: a comparison between the Kellgren and Lawrence system and OARSI atlas criteria. *Knee Surg Sports Traumatol Arthrosc.* 2015;23(12):3532-3539. <http://dx.doi.org/10.1007/s00167-014-3205-0>
 3. Sheehy L, Cooke TDV. Radiographic assessment of leg alignment and grading of knee osteoarthritis: a critical review. *World J Rheumatol.* 2015;5(2):69-81. <http://dx.doi.org/10.5499/wjr.v5.i2.69>
 4. Wright RW. Osteoarthritis classification scales: interobserver reliability and arthroscopic correlation. *J Bone Joint Surg Am.* 2014;96(14):1145-1151. <http://dx.doi.org/10.2106/JBJS.M.00929>
 5. Croft P. An introduction to the *Atlas of Standard Radiographs of Arthritis.* *Rheumatology.* 2005;44(suppl 4):iv42. <http://dx.doi.org/10.1093/rheumatology/kei051>
 6. Riddle DL, Jiranek WA, Hull JR. Validity and reliability of radiographic knee osteoarthritis measures by arthroplasty surgeons. *Orthopedics.* 2013;36(1):e25-e32. <http://dx.doi.org/10.3928/01477447-20121217-14>
 7. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis.* 1957;16(4):494-502. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1006995>
 8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.
-

Received: April 21, 2016 | **Accepted:** November 9, 2016 | **Published:** November 18, 2016