Published in final edited form as:

PM R. 2009 February ; 1(2): 117–126. doi:10.1016/j.pmrj.2008.08.001.

The inter-tester reliability of physical therapists classifying low back pain problems based on the Movement System Impairment classification system

Marcie Harris-Hayes, PT, DPT, OCS and

Assistant Professor, Program in Physical Therapy, Washington University School of Medicine, Campus Box 8502, St. Louis, MO 63108.

Linda R. Van Dillen, PT, PhD

Associate Professor, Program in Physical Therapy, Washington University School of Medicine, Campus Box 8502, St. Louis, MO 63108.

Marcie Harris-Hayes: harrisma@wustl.edu; Linda R. Van Dillen: vandillenl@wustl.edu

Abstract

Objective—Classification of patients with LBP may be important for improving clinical outcomes and research efficiency. The purpose of this study was to examine the inter-tester reliability of two trained physical therapists to classify patients with low back pain (LBP) using the standardized Movement System Impairment (MSI) classification system. The five proposed MSI classifications are based on the most consistent patterns of movement and alignment observed throughout the examination that correlate with the patient's symptom behavior.

Design—Test-retest to assess reliability

Setting—Academic healthcare center outpatient facility

Participants—Thirty subjects with chronic, recurrent LBP (mean age 31.1 ± 12.9 years, 21 F:9 M) were examined independently by two experienced physical therapists.

Methods—Training consisted of self-study of a procedure manual, supervised practice of examination procedures and classification rules and discussion. Subjects were examined independently by each therapist using a test-retest design. Each therapist assigned a LBP classification upon completion of the examination. Both therapists were blinded to the other therapist's findings.

Main Outcome Measures—Inter-tester reliability of therapists classifying the LBP problems was indexed by the percent agreement and kappa coefficient.

Address all correspondence to Dr. Marcie Harris-Hayes, 4444 Forest Park Blvd., Box 8502, Program in Physical Therapy, Washington University School of Medicine, St. Louis, MO 63110; Telephone number: 314 286 1435; FAX: 314 286 1410; E-Mail: harrisma@wustl.edu.

The protocol used for the current study was approved by the Human Studies Committee of Washington University.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

I affirm that I have no financial affiliation (including research funding) or involvement with any commercial organization that has a direct financial interest in any matter included in this manuscript, except as disclosed in an attachment and cited in the manuscript. Any other conflict of interest (ie, personal associations or involvement as a director, officer, or expert witness) is also disclosed in an attachment.

Results—Overall percent agreement on the classification assigned was 83% with kappa =.75 (95% CI =.51 to .99; $P<.0001$).

Conclusion—Inter-tester reliability of classification of patients with LBP using a standardized clinical examination based on the MSI classification system is substantial.

Keywords

Physical therapy; Low back pain; Classification; Spine; Consistency

Introduction

Low back pain (LBP) is a common condition that is associated with significant economic burden.^{1–5} Despite advances in clinical assessment and imaging there is no consistent evidence supporting any one conservative treatment approach for effective management of LBP. Thus, conservative treatment continues to be a challenge. Authors have suggested that the lack of consistent evidence to support conservative treatment may be due to the use of heterogeneous study populations for comparison.^{6, 7} Several investigators have suggested that classification of individuals with LBP into homogeneous subgroups may result in more efficient treatment strategies and improved clinical outcomes.^{8–10}

Classifying homogeneous subgroups of people with LBP may increase the likelihood of responding to specific treatment. Traditionally, classification of LBP has been based on pathoanatomy; however pathoanatomy is purported to be identified in only 10% of patients with LBP.¹¹ Thus, classification based on pathoanatomy may not be the most effective method to guide treatment. Classification based on information collected from the clinical examination may be useful in identifying subgroups and guiding treatment choices.

Although many impairment-based classification systems for LBP have been proposed,^{12–17} only 3 systems were designed to direct rehabilitation and have been studied to some degree. The 3 systems that meet these criteria are the McKenzie (MK) LBP Classification system,¹⁵ the Treatment-Based Classification (TBC) system,¹⁶ and the Movement-System Impairment (MSI) Classification system for LBP.¹⁷ Table 1 provides an overview of the major similarities and differences across the 3 systems.

The primary basis for classification in the MK system is the person's symptom behavior with spinal movements and sustained postures performed within a clinical examination. Specifically, symptoms are assessed with a series of single and repeated spinal movements or prolonged postures in different directions. The purpose of the symptom testing is to identify the pattern of spinal movements and postures that worsen and improve the person's symptoms. Generally, the repeated spinal movements or postures that improve the symptoms (for example repeated extension in standing) are then prescribed as the exercise to manage the LBP.¹⁵

The TBC system was developed in response to limitations noted in the MK system.¹⁶ Similar to the MK examination the TBC examination includes assessment of symptoms with single and repeated spinal movements and sustained postures. Together, symptoms during testing and physical examination signs provide a basis for the person's LBP classification. Treatments based on the person's classification may include stabilization exercise, passive mobilization and manipulation, repeated spinal movements and sustained positions or traction. Instruction in modification of functional activities is general and provided only for the flexion and extension syndromes of the "specific exercise" classification.

Similar to the MK and TBC systems, the MSI system includes descriptions of classifications of LBP based on impairments related to symptoms and mechanical factors identified during a

standardized examination. The MSI examination used to classify is similar to the MK and TBC examinations in that it includes tests of single trunk movements and symptoms with these movements. There are, however, several differences. First, the MSI examination is focused on symptoms produced not only with overt trunk movements (e.g., forward bending) but also with limb movements (e.g., hip rotation in prone).^{18, 19} The initial movement and alignment tests in which the person uses his preferred strategy are referred to as primary tests. Second, as an alternative to performing tests of repeated spinal movements, in the MSI examination symptomatic tests are immediately followed by a secondary test in which the person's preferred alignment or movement strategy is modified. The effect of the secondary test is assessed relative to the symptomatic primary test. Overall, the modifications involve changing a person's strategy by either 1) positioning the lumbar region in a neutral alignment, or 2) restricting lumbar region movement and encouraging movement in other segments (e.g., thoracic region or hip) as the person performs the secondary test.²⁰ Third, examiner judgments of the characteristics of movement focus on amount of motion as well as the relative timing of movements of the spine and proximal limb joints.

The MSI classification of the LBP problem is based on the most consistent pattern of movement and alignment observed throughout the examination and that correlate with the individual's symptom behavior. The MSI classifications are named for the direction(s) of movements and alignments that appear to contribute to the LBP problem. The classifications include lumbar (1) extension, (2) flexion, (3) rotation, (4) extension with rotation, and (5) flexion with rotation.²¹ Studies to validate 3 of the 5 classifications²² and various test items have been reported.^{19, 23–27}

In order for a classification system to be useful, examiners must be able to determine an individual's classification reliably. Thus far, only one study has been published on the ability of examiners to classify LBP problems using the MSI system. Our research group reported on the inter-tester reliability of physical therapists using the standardized examination to classify LBP based on the MSI Model for LBP.²⁸ We reported that the original group of therapists demonstrated moderate²⁹ reliability in classifying the LBP problem.²⁵ The prior investigation of inter-tester reliability, however, included only therapists who were involved in the development of the examination for classifying LBP and had worked together extensively before testing their reliability. In addition, in our previous study, only one examination was performed with each patient while both therapists were present, one examiner and one observer.

The purpose of this study, therefore, was to examine the inter-tester reliability of 2 trained physical therapists to classify individuals with LBP based on the MSI classification system using a test-retest design. One therapist was involved in the development of the standardized examination for classifying. The second therapist involved in the study was not involved in the development process. Examination of the reliability of a new cohort of examiners using a test-retest design is an important step in increasing the generalizability of our previously reported findings. We hypothesized that experienced physical therapists could reliably classify people with LBP into subgroups based on the MSI model.

Methods

Examination

This study was approved by the Human Research Protection Office of *blinded*. A standardized examination, based on the MSI classification system was used to examine and classify subjects with LBP.^{17, 28} The goal of the examination is to identify the direction-specific movement and alignment strategies that consistently reproduce or increase a subject's symptoms. A description of the five MSI subgroups has been published previously.²²

Examiners

Two physical therapists participated in the study. Each therapist had greater than 10 years of experience treating people with musculoskeletal conditions. The first author (MHH) was a board certified clinical specialist in orthopedic physical therapy, who had received instruction in using the MSI system through continuing education and used the concepts of the system in her clinical practice for 7 years. The first author was not involved in the development of the examination. The second author (LVD) had primary responsibility for developing the examination²⁸ used in the current study and has studied properties of the examination and related treatment extensively.^{18–20, 22–25, 28, 30–32} Neither author was examined for reliability in the previous inter-tester reliability study.^{25, 28}

Training

The first author was trained by the second author in the operational definitions for examination items and responses, examination procedures and rules for classification. Training consisted of self-study of an operations manual and practice in the examination procedures with people with and without LBP (N=10). The operations manual included 1) the operational definitions for test items and responses that might be demonstrated by the subject, 2) the construct that each item is proposed to represent, 3) procedures for performing the examination, and 4) explicit rules for classifying the LBP problem. The rules for classification are included in Appendix A. The second author was present during practice to ensure proper performance of the examination procedures and application of the classification rules. The first author practiced using an examination form to record results from the examination and classification decision. There was time allowed for discussion after each training session.

Subjects

Thirty subjects with chronic, recurrent LBP (mean age 31.1 ± 12.9 years, 21 F:9 M) were examined independently by the 2 therapists. Subjects were recruited from the community through newspaper and television advertisements, flyers placed in physician and physical therapy clinics and our University's volunteer registry. Potential subjects contacted the research coordinator in the laboratory where the study was conducted. A detailed description of the study was provided to the potential subject and then he was asked if he would like to participate. If the subject indicated interest, he was then screened more extensively during the telephone interview to determine his eligibility based on the inclusion and exclusion criteria. Subjects between the ages of 18 and 60 who had symptoms related to a LBP problem were eligible to participate. Low back pain symptoms could include pain and paresthesias in the region of the lower back, proximal lower extremity or distal lower extremity.⁷ Subjects were excluded in the case of pregnancy or if they had been previously diagnosed by a physician with one or more of the following conditions: severe kyphosis, scoliosis, spinal stenosis, spinal surgery in the prior 3 months, more than one surgical procedure of the spine, cancer, rheumatoid arthritis, ankylosing spondylitis, or neurological disease. Subjects were also excluded if they were pending spinal surgery or were unable to stand and walk without an assistive device. All subjects read and signed an informed consent statement approved by *Blinded* Human Research Protection Office before participating in the study.

Procedures of Testing and Classification

Subjects were examined independently by each therapist using a test-retest design. Each subject was examined by both therapists on the same day with a 15 minute break between examinations. The order in which the therapists performed the examination was determined by convenience of the therapists' schedules. Examinations were performed in an enclosed treatment room. The therapist waiting to perform the examination was not allowed in the treatment room or surrounding laboratory area.

Prior to the testing session, the subject completed a set of self-report forms that captured data regarding demographics, general health status (SF-36)³³, LBP history and LBP-related functional limitations (Oswestry Low Back Pain Disability Index)³⁴. The first therapist then obtained history information and performed the physical examination. A standardized clinical examination form was used to record findings and the LBP classification.

While the subject rested the first therapist reviewed the self-report and history information with the second therapist. The second therapist then independently performed the physical examination and recorded her findings and LBP classification. Both therapists were blinded to the other therapist's findings and were not allowed to discuss examination procedures during the testing phase of the study. Data forms were collected by a research coordinator after each testing session. A research assistant, independent of the data collection process, entered the data into text files and Systat 10.2 data files.

Statistical Analysis

Descriptive statistics were calculated for demographic, general health status and LBP history variables. Percentage of agreement and a kappa coefficient³⁵ were used to examine the inter-tester reliability of the therapists to classify the subjects with LBP. The 5 possible response categories for classifying a subject's LBP problem were the proposed MSI LBP classifications.

Results

Summary information regarding subject characteristics is provided in Table 2. Thirty subjects (mean age 31.1 ± 12.9 years, 21 F: 9 M) with chronic, recurrent LBP were examined. Eighty percent of the subjects reported symptoms in the back region only⁷, the remaining 20% reported symptoms in the low back and into the lower extremity. Subjects reported minimal LBP-related disability based on the Oswestry Disability Index (13.6 ± 7.5). On average, the subjects reported 3.8 ± 3.4 acute flare ups³⁶ in the previous 12 months.

The first author performed the examination first for 16 of the 30 (53%) examinations. The therapists' responses are summarized in Table 3. Overall percent agreement for the LBP classification assigned was 83% with a kappa value of .75 (95% CI=.51 to .99; $z=6.17$, $P<.0001$). One subject was unable to be classified by either therapist. The therapists disagreed on 5 subjects. One subject was unable to be classified by one of the therapists and was classified as lumbar rotation by the other therapist. Two subjects were classified as lumbar extension with rotation by one therapist and lumbar rotation by the other therapist. Two subjects were classified as lumbar flexion with rotation by one therapist and lumbar rotation by the other therapist.

Discussion

A prerequisite for testing the usefulness of a classification system for LBP is for examiners to be able to reliably classify the proposed LBP problems. In the current study, two experienced physical therapists trained in a standardized examination for classifying LBP based on the MSI model, were able to determine the LBP classifications for a sample of people with non-specific LBP with substantial agreement.²⁹ The current study extends our previous findings^{25, 28} by demonstrating that, with training, a physical therapist not involved in the development of the examination can reliably classify people with LBP. These findings also demonstrate that substantial reliability can be attained using a test-retest design instead of a simultaneous observation design as used previously.²⁵ Such findings extend the generalizability of our initial reliability testing and suggest that the examination could be used by other experienced physical therapists reliably given the appropriate training. Training in this study included self study of a procedure manual, supervised practice of examination procedures and application of the

classification rules to 10 individuals as well as discussion. We believe, however, that learning the specific classification rules was key to attaining a high level of reliability.

We demonstrated similar reliability values as those reported for application of the MK system to classify.^{37, 38} Two recent studies assessing the inter-tester reliability^{37, 38} of the MK system involved physical therapists who had received extensive training and reported a minimum of 5 years of clinical experience using the MK system. All therapists were credentialed in the examination procedures of the MK system. The first study by Kilpikoski et al³⁸ used test-retest methods similar to our study and reported that 2 physical therapists obtained an overall percent agreement for LBP classification of 95% and a kappa value of 0.6. Razmjou et al³⁷ used simultaneous testing to assess reliability of 2 physical therapists and reported agreement of 93% and a kappa value of 0.7.

The reliability reported in the current study is better or similar to that reported for reliability of physical therapists using the TBC system. Direct comparison of the current results to those of the TBC system is not possible, however, due to differences in therapist characteristics.^{39, 40} Heis et al³⁹ examined the reliability of four experienced therapists who were newly trained to apply the TBC classification system. Data from one therapist was not included in the final analysis due to low agreement with the other therapists in the study. The authors reported that the agreement of the three remaining therapists was 55% with a kappa value of 0.46. Fritz et al⁴⁰, also reported on a test-retest design of the reliability of therapists to classify using the TBC system. The therapists had an average of 5.5 years (6 month to 15 years) experience using the TBC system. Agreement of the seven therapists in making a classification decision was 65% with a kappa value of 0.56. Therapist training was not described. Table 4 provides a summary of the results of studies of the reliability of different cohorts of therapists to classify using the three classification systems (MK, TBC MSI),

We have now reported on the inter-tester reliability of physical therapists classifying LBP in two independent samples.^{25, 28} The methods in the current study were more rigorous than our prior work and yet the obtained agreement is higher in the current study than that obtained in our previous study.^{25, 28} The improvement in agreement is likely due to the therapists having more explicit rules for classifying than those available for our original reliability study (Appendix A). The prior study was the first attempt to test the measurement properties of the test items used in the examination, and the primary goal of the study was to examine the ability of therapists to make reliable judgments about individual items from the examination. The rules provided for classifying during the original study were more general than our present rules, and during training less emphasis was placed on learning and applying the rules for assigning a LBP classification than on making judgments about individual test items.^{25, 28} Information obtained from our original reliability study^{25, 28} and subsequent studies^{18–20, 22, 23, 32} have allowed us to develop more specific guidelines for making judgments during individual test items, and to develop more detailed rules for classification. Clarification of criteria for judgments during the examination and development of more specific classification rules likely contributed to our improved therapist agreement in the current study.

Currently, to assign a classification with the MSI system, symptoms must be either produced or increased with some test items during the examination. One subject reported no change in symptoms during either examination. A second subject reported no change in symptoms during the first examination and reported one test as symptom-provoking during the second examination. Following the rules for classifying, both therapists did not assign a classification to the first subject. For the second subject, the first therapist did not assign a classification while the second therapist was able to assign a classification. Thus, a limitation of our current criteria for classification is that an examiner may not be able to classify subjects with a low level of symptom irritability during the examination. After analysis of reliability was completed, the

charts for the two subjects described were examined. In the instances where a classification was not assigned, each therapist recorded what she believed the patient's LBP classification would be based solely on judgments of signs with tests of movement and alignment across the examination. In both subjects, the therapists agreed upon the classification, even though there were little to no tests that evoked symptoms. The criterion of symptom reproduction during the examination, therefore, may represent a limitation in the classification rules. Based on the example from the current study it may be possible that the classification rules could be modified to permit classification based on the signs during tests of movement and alignment made across the examination in the absence of symptom production.

We did not display perfect agreement to classify the LBP problems present in our sample. The therapists disagreed on the classification of five subjects. To determine the nature of our disagreements, we reviewed the data from the examination forms of the subjects for which there was disagreement. The first disagreement is described in the previous paragraph. Two additional classification disagreements were due to the therapists' interpretation of symptoms during individual examination items. Specifically, two subjects described "pressure" in their low back region with a number of the items. One therapist interpreted the "pressure" as the subjects' symptoms; the other therapist did not. Thus, the classification disagreement in these two cases was a result of the therapists' interpretation of symptom behavior. In one subject, the therapists did not agree on the patient's symptom report on a number of test items. The differences could have been a result of subject variability or misinterpretation of symptoms by the examining therapist. On inspection of the examination data from the fifth subject the therapists agreed on the responses to individual items across the examination. One therapist, however, chose a classification inconsistent with the rules. Thus, one therapist misapplied the rules to classify the subject's LBP problem.

We consider the use of a test-retest design in this study to add to the strength of our findings. In our previous work,^{25, 28} both therapists were present during the assessment of each patient. One therapist performed the examination, while the second therapist observed. The simultaneous observation method used in the previous study was intended to remove any variability in patient status or in therapist methods that could affect the results of a test-retest study design. Since the prior study was our first attempt to examine any of the measurement properties of the examination and classification system, the primary question we asked was whether, when the therapists see and hear the same responses could they make the same judgments. The use of the simultaneous observation method, therefore, could have positively affected our prior inter-tester agreement.²⁶ In the current study, each therapist performed the examination independently. Despite possible variability in patient status and variability in methods between the two therapists, our inter-tester reliability was substantial.²⁹

The current study has limitations. First, the extent to which the current sample is representative of all individuals with LBP is not known. Our sample of subjects was recruited from the community through advertisements, flyers placed in physician and physical therapy clinics and a University web-based volunteer registry. The subjects in our sample, therefore, may not represent all individuals with LBP who would present to a medical facility for treatment. The subjects, however, had similar Oswestry scores and pain location and severity as patients who typically are referred to our clinical setting. In addition, our subjects had chronic, recurrent LBP and minimal disability as indexed by the scores on the Oswestry Disability Index. We do not know if we would have similar results in subjects with an acute onset of LBP or with higher levels of disability. Future studies are needed to assess the use of the examination in subjects with acute LBP or higher levels of LBP-related disability.

A second potential limitation is the truncated distribution of the LBP classifications identified in the study sample. There were no subjects classified as lumbar flexion or lumbar extension

in the current sample. Such a finding might also suggest that our study population may not represent all patients with LBP. We do know based on prior data^{22, 41} as well as those of others⁴² that the prevalence of lumbar flexion and lumbar extension problems appears to be less than that of the other proposed classifications. Although the percent agreement was 83%, the skewed distribution of subjects across categories may have contributed to an attenuation of the kappa value.

A third potential limitation is the fact that both examinations for each subject were performed within the same day. We chose to perform both examinations on the same day, however, to ensure stability of subject responses. Stability of the subjects' behavior is an important assumption of a test-retest design so that any differences between test sessions is due to variability in therapist methods and not a result of true change in the subject over time.⁴³ We also examined people within the same day to make the study more feasible for subjects to participate. A potential disadvantage of repeated testing in the same day is that subjects' symptoms could be increased during the second examination compared to the first examination. Any differences in subjects between the two sessions, however, did not substantially affect our reliability as evidenced by the kappa value ($k=.75$) obtained.

Finally, the generalizability of our findings to other examiners may still be somewhat limited. Both therapists were experienced in treating musculoskeletal pain problems and had practiced applying the concepts of the MSI model for LBP to patients. The first author had used the examination and treatment principles in her clinical practice across 7 years. The second author had primary responsibility in developing the examination and used the procedures extensively in prior studies. We do not know if we would find similar reliability in examiners with less clinical experience or less experience applying the principles of the model that is the basis for the MSI classification system. Our primary purpose with the current study, however, was to examine what therapists' reliability to classify would be when we used a more rigorous study design (test-retest design) and when someone who was not involved in the original development of the examination was tested. The current study suggests that the reliability to classify people with LBP under more stringent conditions is actually better than that attained in our earlier reliability study. An appropriate follow-up to the current work would be to examine the inter-tester reliability of novice, but trained examiners. Such work is currently underway. After a two day instructional course, 13 examiners with no experience to moderate experience with the MSI classification system classified written cases of data from people with LBP. Agreement among therapists was excellent with an overall kappa of 0.81 (CI: 0.78–0.83, $p<0.01$) (unpublished data).

Acknowledgments

Funded by the US National Institutes of Health (NIH), grant number: 52833.

Reference List

1. Jette AM. Physical therapy episodes of care for patients with low back pain. *Phys Ther* 1994 February; 74(2):101–10. [PubMed: 8290616]
2. Lawrence RC, Helmick CG, Arnett FC, et al. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. *Arthritis Rheum* 1998 May;41(5):778–99. [PubMed: 9588729]
3. Walker BF. The prevalence of low back pain: a systematic review of the literature from 1966 to 1998. *J Spinal Disord* 2000 June;13(3):205–17. [PubMed: 10872758]
4. Rubin DI. Epidemiology and risk factors for spine pain. [Review] [92 refs]. *Neurol Clin* 2007 May;25 (2):353–71. [PubMed: 17445733]

5. Maniadas N, Gray A. The economic burden of back pain in the UK. *Pain* 2000 January;84(1):95–103. [PubMed: 10601677]
6. Leboeuf-Yde C, Lauristen JM, Lauristen T. Why has the search for causes of low back pain largely been nonconclusive? *Spine* 1997;22:877–81. [PubMed: 9127921]
7. Spitzer WO. Diagnosis of the problem (the problem of diagnosis). In: Scientific approach to the assessment and measurement of activity-related spinal disorders: a monograph for clinicians' report of the Quebec task force on spinal disorders. *Spine* 1987;12(suppl S16):S16–S21.
8. Sahrman SA. Diagnosis by the physical therapist--a prerequisite for treatment. A special communication. *Phys Ther* 1988 November;68(11):1703–6. [PubMed: 3054944]
9. Rose SJ. Musing on diagnosis [editorial]. *Phys Ther* 1988;68:1665. [PubMed: 3186790]
10. Borkan JM, Cherkin DC. An agenda for primary care research on low back pain. *Spine* 1996;24:2880–4. [PubMed: 9112712]
11. Abenhaim L, Rossignol M, Gobeille D, Bonvalot Y, Fines P, Scott S. The Prognostic Consequences in the Making of the Initial Medical Diagnosis of Work-Related Back Injuries. *Spine* 1995 April;20(7):791–5. [PubMed: 7701392]
12. Wilson L, Hall H, McIntosh G, Melles T. Intertester reliability of a low back pain classification system. *Spine* 1999 February;24(3):248–54. [PubMed: 10025019]
13. Dankaerts W, O'sullivan PB, Burnett AF, Straker LM. The use of a mechanism-based classification system to evaluate and direct management of a patient with non-specific chronic low back pain and motor control impairment--a case report. *Man Ther* 2007 May;12(2):181–91. [PubMed: 16877024]
14. Bertilson BC, Bring J, Sjoblom A, Sundell K, Streder LE. Inter-examiner reliability in the assessment of low back pain (LBP) using the Kirkaldy-Willis classification (KWC). *Eur Spine J* 2006 November;15(11):1695–703. [PubMed: 16435110]
15. McKenzie, RA. *The Lumbar Spine: Mechanical diagnosis and Therapy*. Waikanae, New Zealand: Spinal publications Ltd; 1998.
16. Delitto A, Erhard RE, Bowling RW. A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative treatment. *Phys Ther* 1995;75:470–89. [PubMed: 7770494]
17. Sahrman, SA. *Diagnosis and Treatment on Movement Impairment Syndromes*. St. Louis: Mosby, Inc; 2002.
18. Van Dillen LR, Sahrman SA, Norton BJ, et al. Effect of active limb movements on symptoms in patients with low back pain. *Journal of Orthopaedic & Sports Phys Ther* 2001;31(8):402–13.
19. Van Dillen LR, Gombatto SP, Collins DR, Engsberg JR, Sahrman SA. Symmetry of timing of hip and lumbopelvic rotation motion in 2 different subgroups of people with low back pain. *Arch Phys Med Rehabil* 2007 March;88(3):351–60. [PubMed: 17321829]
20. Van Dillen LR, Sahrman SA, Norton BJ, Caldwell CA, McDonnell MK, Bloom N. The effect of modifying patient-preferred spinal movement and alignment during symptom testing in patients with low back pain: a preliminary report. *Arch Phys Med Rehabil* 2003 March;84(3):313–22. [PubMed: 12638097]
21. Van Dillen Linda R, Maluf Katrina S, Sahrman Shirley A. Further examination of modifying patient-preferred movement and alignment strategies in patients with low back pain during symptomatic tests. *Man Ther*. 2007in press
22. Van Dillen LR, Sahrman SA, Norton BJ, et al. Movement system impairment-based categories for low back pain: Stage I validation. *J Orthop Sports Phys Ther* 2003 March;33(3):126–42. [PubMed: 12683688]
23. Gombatto SP, Collins DR, Sahrman SA, Engsberg JR, Van Dillen LR. Patterns of lumbar region movement during trunk lateral bending in 2 subgroups of people with low back pain. *Phys Ther* 2007 April;87(4):441–54. [PubMed: 17374634]
24. Gombatto SP, Collins DR, Sahrman SA, Engsberg JR, Van Dillen LR. Gender differences in pattern of hip and lumbopelvic rotation in people with low back pain. *Clin Biomech* 2006 March;21(3):263–71.
25. Norton BJ, Sahrman SA, Van Dillen FL. Differences in measurements of lumbar curvature related to gender and low back pain. *J Orthop Sports Phys Ther* 2004 September;34(9):524–34. [PubMed: 15493520]

26. Scholtes, S.; Gombatto, S.; Collins, D.; Sahrman, S.; Van Dillen, L. Differences in lumbopelvic movement during knee flexion in prone between people with and without low back pain. In 6th Interdisciplinary World Congress on Low Back and Pelvic Pain; November 7–10; 2007. p. 343–4.
27. Scholtes SA, Gombatto SP, Van Dillen LR. Gender differences in timing of lumbopelvic movement during the clinical test of active knee flexion. *J Orthop Sports Phys Ther* 2008;38(1):A69.
28. Van Dillen LR, Sahrman SA, Norton BJ, et al. Reliability of physical examination items used for classification of patients with low back pain. *Phys Ther* 1998 September;78(9):979–88. [PubMed: 9736895]
29. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74. [PubMed: 843571]
30. Harris-Hayes M, Van Dillen LR, Sahrman SA. Classification, treatment and outcomes of a patient with lumbar extension syndrome. *Physiother Theory Pract* 2005 July;21(3):181–96. [PubMed: 16389699]
31. Maluf KS, Sahrman SA, Van Dillen LR. Use of a classification system to guide nonsurgical management of a patient with chronic low back pain. *Phys Ther* 2000;80:1097–111. [PubMed: 11046197]
32. Van Dillen LR, Sahrman SA, Caldwell CA, McDonnell MK, Bloom N, Norton BJ. Trunk rotation-related impairments in people with low back pain who participated in 2 different types of leisure activities: a secondary analysis. *J Orthop Sports Phys Ther* 2006 February;36(2):58–71. [PubMed: 16494073]
33. Ware JE, Sherbourne CD. The Mos 36-Item Short-Form Health Survey (Sf-36).1. Conceptual-Framework and Item Selection. *Med Care* 1992 June;30(6):473–83. [PubMed: 1593914]
34. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66:271–3. [PubMed: 6450426]
35. Cohen J. Coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960;20:37–46.
36. Von Korff M. Studying the natural history of back pain. *Spine* 1994 September 15;19(18Suppl):Suppl-2046S.
37. Razmjou H, Kramer JF, Yamada R. Intertester reliability of the McKenzie evaluation in assessing patients with mechanical low-back pain.[see comment]. *J Orthop Sports Phys Ther* 2000 July;30(7):368–83. [PubMed: 10907894]discussion 384–9
38. Kilpikoski S, Airaksinen O, Kankaanpää M, Leminen P, Videman T, Alen M. Interexaminer reliability of low back pain assessment using the McKenzie method. *Spine* 2002 April;27(8):E207–14. [PubMed: 11935120]
39. Heiss DG, Fitch DS, Fritz JM, Sanchez WJ, Roberts KE, Buford JA. The interrater reliability among physical therapists newly trained in a classification system for acute low back pain. *J Orthop Sports Phys Ther* 2004 August;34(8):430–9. [PubMed: 15373006]
40. Fritz JM, George S. The use of a classification approach to identify subgroups of patients with acute low back pain. Interrater reliability and short-term treatment outcomes. *Spine* 2000 January;25(1):106–14. [PubMed: 10647168]
41. Van Dillen Linda R, Sahrman Shirley A. Outcomes of classification-directed intervention in people with chronic or recurrent low back pain. [abstract]. *J Orthop Sports Phys Ther* 2006;36:A61.
42. Trudelle-Jackson E, Sarvaiya-Shah SA, Wang SS. Interrater reliability of the movement impairment based classification system for lumbar spine syndromes in patients with chronic low back pain [abstract]. *J Orthop Sports Phys Ther*. 2008in press
43. Stratford PW. Getting more from the literature: estimating the standard error of measurement from reliability studies. *Physiother Can* 2004;56(1):27–30.
44. McGill SM, Kippers V. Transfer of loads between lumbar tissues during the flexion-relaxation phenomenon.[see comment]. *Spine* 1994;19:2190–6. [PubMed: 7809753]

Appendix

Appendix A Rules for Classification into Low Back Pain Subgroups

1. Overall guidelines for decision making:
 - a. Greater weight is given to symptom behavior with tests than signs with tests.
 - b. Greater weight is given to tests in which the person reports
 - i. An increase in symptoms with a primary test that also has an associated sign and a decrease or elimination of symptoms with the associated secondary test.
 - ii. An increase in symptoms with a primary test and a decrease or elimination of symptoms with the associated secondary test.
 - iii. A decrease in symptoms with a primary test associated with the opposite direction of movement or alignment. For example, a person may report an increase in symptoms with trunk extension in standing and a decrease in symptoms with trunk flexion in standing.
2. Lumbar extension:
 - a. Symptom behavior
 - i. The person reports an increase in symptoms with primary tests related to extension movements and alignments.
 - ii. The person reports a decrease or elimination of symptoms with associated secondary tests for the tests related to extension movements and alignments.
 - iii. The person *may* report a decrease or elimination of symptoms with primary tests associated with flexion.
 - iv. Overall, the person does *not* report an increase in symptoms with primary tests related to flexion or rotation movements and alignments.
 - a. The person may report an increase in symptoms with one flexion test and still be classified as Lumbar extension. An increase in symptoms with a flexion test is proposed to occur in circumstances in which one or more lumbar region segments shear.⁴⁴ Thus, flexion of the lumbar region is not contributing to the increase in symptoms. An increase in symptoms due to shearing is thought to occur most often with the test of forward bending in standing.
 - b. Signs
 - i. The person may display signs associated with tests of extension, in particular, an alignment of increased lumbar extension in different positions.
3. Lumbar extension with rotation:

a. Symptom behavior

- i.** The person reports an increase in symptoms with primary tests related to extension, rotation or combined extension and rotation movements and alignments.
- ii.** The person reports a decrease or elimination of symptoms with associated secondary tests for the tests related to extension, rotation or combined extension and rotation movements and alignments.
- iii.** The person *may* report a decrease or elimination of symptoms with primary tests associated with flexion.
- iv.** Overall, the person does *not* report an increase in symptoms with primary tests related to flexion or combined flexion and rotation movements and alignments.
 - a.** The person may report an increase in symptoms with one flexion test and still be classified as Lumbar extension with rotation. An increase in symptoms with a flexion test is proposed to occur in circumstances in which one or more lumbar region segments shear.⁴⁴ Thus, flexion of the lumbar region is not contributing to the increase in symptoms. An increase in symptoms due to shearing is thought to occur most often with the test of forward bending in standing.
- v.** The person must report an increase in symptoms with at least one extension test and one rotation test or one combined extension and rotation test to be classified as a Lumbar extension with rotation. Please see 3.b.iii for the exception to this rule.

b. Signs

- i.** The person may display signs with tests associated with extension, rotation or combined extension with rotation. Some of these signs will be associated with an increase in symptoms.
- ii.** The person will tend to display an asymmetry in the rotation-related signs.
- iii.** In some instances the rotation signs will be prevalent during the examination but not be associated with an increase in symptoms. In these circumstances the rotation component may still be important to the LBP problem. Therefore, if the person displays one third (N=6) or more of the rotation signs across the examination, but these signs are not associated with any increase in symptoms, the examiner may consider the signs as important in the decision-making process. The classification of Lumbar extension with rotation may be assigned.

4. Lumbar rotation:**a. Symptom behavior**

- i. The person reports an increase in symptoms with primary tests related to rotation, flexion and extension movements and alignments.
 - a. The person must report an increase in symptoms with at least 2 primary tests related to flexion movements or alignments to be classified as Lumbar rotation. This criterion is required because it is proposed that a person may actually have an increase in symptoms with some of the tests related to flexion as a result of shearing of lumbar region segments.⁴⁴ An increase in symptoms due to shearing is thought to occur most often with the test of forward bending in standing.
- ii. The person reports a decrease or elimination of symptoms with associated secondary tests for the tests related to rotation, flexion, and extension.

b. Signs

- i. The person *may* display signs associated with tests of rotation, flexion, extension, combined flexion and rotation or combined extension and rotation. The predominance of signs will be rotation-related.
- ii. Some of the signs will be associated with an increase in symptoms.
- iii. The person will tend to display symmetry of the rotation-related signs.

5. Lumbar flexion with rotation:

a. Symptom behavior

- i. The person reports an increase in symptoms with primary tests related to flexion, rotation or combined flexion and rotation movements and alignments.
- ii. The person reports a decrease or elimination of symptoms with the associated secondary tests for the tests related to flexion, rotation or combined flexion and rotation movements and alignments.
- iii. Overall, the person does *not* report an increase in symptoms with primary tests related to extension, rotation or combined extension and rotation movements and alignments.
- iv. The person must report an increase in symptoms with at least two flexion tests and one rotation test or one combined rotation with flexion test to be classified as a Lumbar flexion with rotation. The more tests of flexion, rotation or combined flexion and rotation that increase the person's symptoms, the more likely the classification is Lumbar flexion with rotation. Please see 5.b.iii for the exception to this rule.

b. Signs

- i.** The person will display signs associated with tests of flexion, rotation or combined flexion and rotation movements and alignments. Some of these signs will be associated with an increase in symptoms.
- ii.** The person will tend to display an asymmetry in the rotation-related signs.
- iii.** In some instances the rotation signs will be prevalent during the examination but not be associated with an increase in symptoms. In these circumstances the rotation component may still be important to the LBP problem. Therefore, if the person displays one third (N=6) or more of the rotation signs across the examination, but these signs are not associated with any increase in symptoms, the examiner may consider the signs as important in the decision-making process. The classification of Lumbar flexion with rotation may be assigned.

6. Lumbar flexion:**a. Symptom behavior**

- i.** The person reports an increase in symptoms with primary tests related to flexion movements and alignments.
- ii.** The person reports a decrease or elimination of symptoms with the associated secondary tests for the tests related to flexion movements and alignments.
- iii.** The person *may* report a decrease or elimination of symptoms with primary tests associated with extension.
- iv.** Overall, the person does not report an increase in symptoms with primary tests related to extension, rotation or combined extension and rotation movements and alignments.

b. Signs

- i.** The person may display signs associated with flexion. In particular, the person may display an alignment of lumbar region flexion in quadruped and early lumbar region flexion with rocking back in quadruped.

Table 1

Comparison of features of three classification systems for LBP.

Feature	McKenzie	Treatment-Based Classification (Stage I)	Movement System Impairments
Conceptual Model	LBP is result of varying degrees of mechanical deformation of periarticular tissue &/or disc with prolonged postures & repeated movements of the spine. Majority of LBP problems are proposed to be associated with lumbar flexion. ¹⁵	Basis for LBP is same as MK system for specific exercise, mobilization, & traction classifications. LBP is result of excessive movement of lumbar spine for the immobilization classification. Majority of problems in acute, work-related LBP reported to be in specific exercise classification (lumbar flexion & lumbar extension) & immobilization classification. ⁴³	LBP is result of repeated movements and sustained postures of spine in <i>specific direction(s)</i> resulting in strategies. Continual use of strategies contributes to impairments, accelerated cumulative tissue stress, microtrauma, LBP, macrotrauma. Majority of chronic LBP problems proposed to be associated with lumbar extension &/or rotation. ¹⁷
System Specifics	3 classifications: postural, dysfunction, derangement	4 classifications: specific exercise (flexion, extension, lateral shift), mobilization (sacroiliac & lumbar), traction, immobilization	5 classifications: extension, flexion, rotation, rotation with flexion, rotation with extension
	Named for proposed tissue impairment contributing to LBP, e.g. disc derangement.	Named for treatment to be administered, e.g., traction.	Named for <i>specific direction(s)</i> of movement and alignment contributing to LBP problem, e.g., lumbar rotation with extension.
	Exam includes 1) tests of symptoms with single & repeated end-range spine movements & sustained end-range spine postures, & 2) signs of spine and pelvic movements and alignments.	Exam includes 1) tests of symptoms with single & repeated end-range spine movements & sustained end-range spine postures, & 2) signs of spine and pelvic movements and alignments.	Exam includes 1) primary tests of symptoms with movements(trunk & limb) in 7 different positions; Symptomatic tests followed by secondary or modified test to eliminate symptoms, & 2) signs of spine & pelvic movements & alignments with trunk & limb movements.
Treatment Guidelines	<u>Derangement, Dysfunction</u> : passive (mobilization or manipulation) or active repeated end-range spinal movement; postural instruction sitting, standing, lying; <u>Posture</u> : postural instruction sitting, standing, lying	<u>Specific Exercise, Mobilization, Traction</u> : passive (mobilization, manipulation, traction) or active repeated end-range spinal movement; postural instruction sitting, standing, lying; <u>Immobilization</u> : stabilization with exercise &/or external support	<u>All Categories</u> : education in tissue injury & healing; instruction in modification of <i>direction-specific</i> movement and alignment strategies with symptomatic functional activities; active exercise to modify <i>direction-specific</i> strategies used with symptomatic impairment tests

TABLE 2

Characteristics of study sample. N = 30

Variable	Values
Male (%)	9 (30)
Mean age in years (SD)	31.1 (12.9)
Mean body mass index in kg/m ² (SD)	23.1 (2.2)
Mean SF-36 subscales (SD)	
physical functioning	.91 (.08)
role-physical	.75 (.32)
bodily pain	.65 (.15)
general health	.80 (.13)
vitality	.58 (.19)
social functioning	.87 (.18)
role-emotional	.86 (.32)
mental health	.82 (.10)
Mean Symptom Intensity, ^a 0–10, (SD)	
Current during standing	1.6 (2.0)
Average over prior 7 days	2.7 (1.6)
Worst over prior 7 days	4.4 (1.6)
Mean Oswestry Disability Index, 0–100% (SD) ^b	13.6 (7.5)
Location of symptoms ^c	
low back only (%)	24 (80)
low back and proximal LE (%)	4 (13)
low back and distal LE (%)	1 (3)
low back, proximal and distal LE (%)	1 (3)
Mean number of acute flare ups in prior 12 months, (SD) ^d	3.8 (3.4)
Number (%) in flare-up day of examination	2 (7)

^aVerbal estimate of symptom intensity on a numerical scale (0, no symptoms present; 10, worst imaginable symptoms).⁴⁸

^bOswestry Low Back Pain Disability Index³⁷ is a disease-specific measure that represents the degree of disability as a percentage (0%, no disability; 100% maximal disability).

^cDefinitions from Quebec Task Force on Spinal Disorders.²⁴ Low back, area extending from T12 to gluteal fold; proximal LE, area extending from the gluteal fold to knee; distal LE, area extending from knee to foot.

^dFlare-up is defined as a phase of low back pain that is superimposed on a recurrent or chronic course. During a flare-up, the LBP is noticeably more severe than usual for the subject.³⁹

TABLE 3
Inter-tester agreement, observed and (expected) between physical therapists regarding classification assignments.

Classification by Therapist one	Classification by Therapist two					
	Extension with Rotation	Extension with Rotation	Rotation	Flexion with Rotation	Unable to Classify ^{ab}	Total
	Extension with Rotation	10 (4.1)	1 (5.1)	0 (1.5)	0 (0.4)	11
	Rotation	1 (4.1)	10 (5.1)	0 (1.5)	0 (0.4)	11
	Flexion with Rotation	0 (2.2)	2 (2.8)	4 (0.8)	0 (0.2)	6
	Unable to Classify ^a	0 (0.7)	1 (0.9)	0 (0.3)	1 (0.1)	2
Total	11	14	4	1	30 ^b	

^aUnable to classify because symptoms were not reproduced during the examination

^bPercent agreement: 25/30 (83%); $\kappa = 0.75$

TABLE 4
Inter-tester reliability of impairment based classification systems.

Study	Classification ^a system	Percent agreement	Kappa	Subjects (N)	Examiner characteristics
Harris-Hayes and Van Dillen, Current	MSI	83	0.75	30	Examiners: 2 physical therapists Clinical experience: greater than 7 years Experience with system: greater than 7 years Training: consisted of attendance of continuing education course, self study of procedure manual, and supervised practice of exam procedures.
Heis et al, 200442	TBC	55	0.46	45	Examiners: 3 physical therapists ^b Clinical experience: greater than 12 years Experience with system: no prior experience Training: consisted of reading an article ¹⁶ describing the system and a 1 day training course that consisted of lecture, discussion, hands-on laboratory, and case study sessions.
Fritz et al, 200043	TBC	65	0.56	43	Examiners: 7 physical therapists Clinical experience: not reported Experience with system: average of 5.5 years Training: not reported.
Kilpikoski et al, 200241	MK	95	0.6	39	Examiners: 2 physical therapists Clinical experience: not reported Experience with system: average of 5 years Training: both held a MK diploma and were credentialed examiners in the MK system.
Razmjou et al, 200040	MK	93%	0.7	45	Examiners: 2 physical therapists Clinical Experience: greater than 12 years Experience with system: not reported Training: One therapist held a MK diploma and one therapist was a faculty member of the MK International Institute.

^a MSI: Movement-System Impairment Classification system
TBC: Treatment-Based Classification system¹⁶
MK: McKenzie LBP Classification system.¹⁵

^b Four physical therapists participated, however data from one therapist was not included in the final analysis due to low agreement with the other therapists in the study.