Published in final edited form as:

Man Ther. 2013 February ; 18(1): 35–40. doi:10.1016/j.math.2012.06.008.

Reliability of novice raters in using the movement system impairment approach to classify people with low back pain

Sharon M. Henry^{a,*}, Linda R. Van Dillen^b, Andrea R. Trombley^c, Justine M. Dee^c, and Janice Y. Bunn^d

^a Department of Rehabilitation Science, 305 Rowell Building, University of Vermont, Burlington, VT 05401, USA

^b Program in Physical Therapy and Department of Orthopaedic Surgery, Washington University, School of Medicine, St. Louis, MO 63108, USA

^c Department of Rehabilitation and Movement Science, University of Vermont, Burlington, VT, USA

^d Department of Mathematics and Statistics, University of Vermont, Burlington, VT, USA

Abstract

Observational cross sectional study.

To examine the inter-rater reliability of novice raters in using the Movement System Impairment (MSI) approach system and to explore the patterns of disagreement in classification errors. The inter-rater reliability of individual tests items used in the MSI approach is moderate to good; however, the reliability of the classification algorithm has been tested only preliminarily.

Using previously recorded patient data ($n = 21$), 13 novice raters classified patients according to the MSI schema. The overall percent agreement using the kappa statistic as well as the agreement/disagreement among pair-wise comparisons in classification assignments were examined.

There was an overall 87.4% agreement in the pairs of classification judgments with a kappa coefficient of 0.81 (95% CI: 0.79, 0.83). Raters were most likely to agree on the classification of Flexion (100%) and least likely to agree on the classification of Rotation (84%).

The MSI classification algorithm can be learned by novice users and with training, their inter-rater reliability in applying the algorithm for classification judgments is good and similar to that reported in other studies. However, some degree of error persists in the classification decision-making associated with the MSI system, in particular for the Rotation category.

Keywords

Classification of low back pain; Reliability; Subgrouping of low back pain

1. Introduction

To date, no treatment strategy has been consistently effective in reducing the typical long-term symptoms, functional limitations and disability associated with low back pain (LBP). The limited beneficial effects of conservative treatments for LBP may be due to the lack of a

pathoanatomical-specific diagnosis (Mooney, 1989; Nordin et al., 1998; Kent and Keating, 2005). In the absence of a specific pathoanatomical diagnosis, researchers have proposed that people with LBP represent a heterogeneous group and should be classified based on key clinical symptoms and multidimensional features of their LBP presentation in order to better direct treatment (Moffroid et al., 1994; Atlas et al., 1996; Borkan and Cherkin, 1996; Croft, 1997; Billis et al., 2007; Fritz et al., 2007; Dankaerts and O'Sullivan, 2011). Logically, homogeneous subgroups would have a higher likelihood of responding to matched treatment approaches. Such classification systems could be useful in prognosis and intervention planning, making classification systems a top LBP research priority (Spitzer et al., 1987; Borkan and Cherkin, 1996; Croft, 1997).

Numerous attempts to develop classification systems for patients with LBP have produced schemas that are biophysically based, psychosocially based or both (for review see Billis et al., 2007). The Movement System Impairment (MSI) approach (Sahrmann, 2002) is one classification system that is biophysically based and shows promise because it directs intervention and assists prognosis based on patient-specific criteria that cluster patients into homogeneous subgroups. The theoretical model upon which the MSI approach is built asserts that the maintenance of precise motion at each joint segment ensures optimal function, which is achieved through periodic movement and variability in the direction of movement, rather than through repeated movement in one direction. Loss of movement precision is proposed to contribute to LBP when a person exhibits the tendency to move the lumbar spine in a specific direction when repositioning the trunk or limbs (Sahrmann, 2002). Using this framework, the standardized MSI exam attempts to identify consistent patterns of 1) preferred trunk movements, 2) preferred trunk alignments and/or 3) timing of lumbopelvic segment movements relative to proximal joints during standardized tests that are associated with the patient's low back symptoms. The identification of these movement patterns allows patients to be classified into one of five subgroups of LBP named for the specific directions of trunk movements and alignments (flexion, extension, and rotation or combinations) that are associated with an increase in the patient's LBP. The five subgroups are: 1) Lumbar Rotation, 2) Lumbar Extension, 3) lumbar Flexion, 4) Lumbar Rotation with Extension, and 5) Lumbar Rotation with Flexion (henceforth, these subgroups will be referred to without including the word 'Lumbar').

During the MSI exam, individual tests that increase the patient's symptoms are immediately followed by a modified test in which the patient's preferred trunk alignment or movement pattern is altered and the effect is assessed relative to the initial symptomatic test. The modified second test involves changing a patient's direction specific pattern used during a previously symptomatic test by either 1) positioning the lumbar spine in a neutral alignment, or 2) restricting lumbar spine movement and encouraging movement in other segments (e.g., thoracic spine or hip). For 82% of these secondary tests, the majority of people with LBP report a decrease or elimination of LBP symptoms with the modified tests, which serves as a confirmation that the direction3 specific trunk alignment and movement(s) contribute to the patient's LBP (Van Dillen et al., 2003a; Van Dillen et al., 2009). Treatment effectiveness studies for the MSI classification system using case based reports (Maluf et al., 2000; Harris Hayes et al., 2005; Van Dillen et al., 2005) demonstrate a superior improvement (1 year) compared to other classification systems (McKenzie and May, 2003) and treatment approaches (Mannion et al., 1999; Koumantakis et al., 2005). However, in a recent trial (Henry et al., 2012), patients receiving MSI-directed treatment did not demonstrate superior outcomes post-treatment ($p = 0.91$) compared to subjects who received Treatment Based Classification-directed stabilization exercises (Fritz et al., 2007). Thus, it is undetermined if MSI treatment leads to superior longer term outcomes (i.e., one and two years) in a clinical trial. Furthermore, it has yet to be determined whether or not all patients with LBP can be classified using the MSI approach.

In order to be useful for research or clinical practice, a classification system must demonstrate certain characteristics related to reliability, feasibility, generalizability, and various aspects of validity (Buchbinder et al., 1996; Riddle, 1998). Several of these characteristics have been investigated with respect to the MSI approach. The construct validity of three MSI categories (rotation with extension, rotation, and extension) has been examined with data from 188 patients with LBP through a principle components analysis that showed three factors of impairment related to each of these three MSI categories (Van Dillen et al., 2003b). The inter-rater reliability of the classification judgments made using the MSI approach are moderate to good with an overall percent agreement on the classification assigned ranging from 75% to 83% with kappa values ranging 0.61–0.75 (Van Dillen et al., 1998; Trudelle-Jackson et al., 2008; Harris-Hayes and Van Dillen, 2009).

Still, there is some degree of error that persists in the classification decision making associated with the MSI approach. One potential source of this error that has not been previously examined is the mutual exclusivity of the classification categories.

Ideally any individual with LBP should fit one and only one classification category. The complex and multidimensional clinical presentation of patients with LBP, however, makes some degree of overlap between categories seem inevitable. Identifying which categories overlap most commonly may assist in refining the decision-making algorithm and ultimately improve the usefulness of the MSI approach. Thus, the purpose of this study was to examine the inter-rater reliability in applying the classification criteria of the MSI approach, and to explore the pattern of disagreements that existed to determine which categories were most difficult to distinguish.

2. Methods

2.1. Subjects

The subjects in this study were 13 raters, 12 of whom were licensed physical therapists, who were instructed to use a published algorithm (Harris-Hayes and Van Dillen, 2009) to determine patient classification according to the MSI schema using previously recorded patient data. The raters completed a questionnaire about their education level, credentials, physical therapy experience and familiarity with the MSI approach (Table 1). All subjects were informed about the experimental protocol and the potential risks of the study and gave written consent prior to their participation. The protocol and consent form were previously approved by the Institutional Review Board at the University of Vermont.

2.2. Materials and patient data collection

The patient data used in the study were taken from previous studies that were designed to identify sub-groupings of patients with LBP (NIH/K01 HD01226-05 and Physical Therapy Foundation Grant to LVD). These studies were approved by the Institutional Review Board of Washington University and all subjects provided written consent. From these studies, 21 cases were selected, ensuring a balanced representation of all five MSI categories, for inclusion in the reliability study on which we are now reporting.

The clinical data collected from the 21 subjects with LBP included self-report questionnaires and a physical examination; these procedures provided the data necessary to classify the subject using the decision-making algorithm (Harris-Hayes and Van Dillen, 2009). The questionnaires administered to the patients included the Modified Oswestry Questionnaire (Fritz and Irrgang, 2001) and an 11-point numeric pain rating scale (Jensen et al., 1994). Demographic information collected included age, sex, prior history of LBP, aggravating and relieving factors, and the duration and location (low back only, below the buttock or below the knee) of current symptoms. The physical examination conducted by licensed physical

therapists included standardized tests to quantify trunk alignment and movement patterns that correlated with the patient's symptoms. The direction-specific trunk alignment and movement impairments across the clinical exam were then used by the therapist to classify the patient with LBP into one of five groups by using an established set of MSI classification rules (Harris-Hayes and Van Dillen, 2009).

2.3. Procedures

Approximately six days of training were provided for the 13 UVM raters during which they were oriented to the goals of the study, the paperwork, the MSI approach and the decision making algorithm. Several practice cases (different than those used for this study) were used and discussed to familiarize the raters with the data and the algorithm. Following that training, each of the 13 raters were given 21 clinical examination forms that included only the subjects' history and physical exam data as well as a recording form for recording the classification choice. Working independently, each rater used the written physical exam data to assign one of five possible MSI classification categories (Rotation, Extension, Flexion, Rotation with Extension, and Rotation with Flexion) to each of the 21 cases using the published classification rules (Harris-Hayes and Van Dillen, 2009). Only one classification judgment was permitted per subject and each rater was instructed to submit a classification judgment for each of the 21 subjects.

In order to examine which classification judgments were most difficult to distinguish from another and to examine inter-rater reliability of the MSI approach, results from each of the 13 raters from the University of Vermont (UVM) were compared to the majority ratings as well as to an expert MSI rater (LVD) who had previously rated each of the 21 cases.

2.4. Data analysis

Descriptive characteristics were examined for the 13 raters and the 21 subjects with LBP. To assess inter-rater reliability of the MSI approach, the distribution of classification judgments were computed and compared among the 13 UVM raters. For each subject with LBP included in the analysis, we compared the judgment of a given rater with the judgments of each other rater and categorized each pair of judgments as "in agreement" (i.e., the same classification judgment for both raters) or "in disagreement" (i.e., different classification judgments for each rater). This was repeated for all of the raters, such that all possible combinations of two raters were examined. The overall percent agreement was calculated as the percent of pairs of judgments in which both raters were "in agreement" compared to the total number of possible rater combinations. Because overall percent agreement is limited by the possibility of classification agreement simply due to chance, additional measures of inter-rater reliability must be employed. One such measure is the kappa coefficient, which measures the proportion of agreement among raters after chance agreement has been removed. An unweighted kappa coefficient examining overall agreement of LBP classification into six categories (the five MSI subgroups and an additional category of "unclassifiable"), as well as kappa coefficients for each category along with 95% confidence intervals, were computed (King, 2004). To explore the nature of the disagreements in the classification decisions, we also examined the percentage of raters who agreed or disagreed with the classifications made by the majority of raters and the classification made by the expert rater. Finally, we categorized each pair of classification judgments that were "in disagreement" into possible categories of disagreement based on the classification judgment of each member of the pair of raters (see Table 6 for list of 15 possible categories of disagreement). The frequency of each disagreement category was calculated.

3. Results

The descriptive characteristics of the 13 raters and the 21 subjects with LBP are provided in Tables 1 and 2, respectively. Of the 21 cases, the expert (LVD) determined the classification of cases as follows: four cases were Rotation; five cases were Extension; three were Flexion; five were Rotation with Extension; three were Rotation with Flexion; and one case could not be classified (Table 3).

3.1. Agreement among raters

Of the possible 273 classifications (21 cases $\times$ 13 raters), there were a total of 269 classification judgments available for analysis and thus 1590 pair-wise comparisons, using each pair only once, among the 13 raters' chosen classification. Overall, 16.0% of the classification judgments were for Flexion; 21.6% for Extension; 18.2% for Rotation with Flexion category; 26.0% for Rotation with Extension; 17.5% for the Rotation category; and 0.7% were unclassifiable (Table 3).

There was an overall 87.4% agreement in the pairs of classification judgments with a kappa coefficient of 0.81 (95% CI: 0.79, 0.83) (Table 4). There was an overall agreement of 90.0% agreement between the UVM raters and the expert (LVD). When examining rater agreement by a particular classification category, the kappa coefficient was 0.90 (95% CI: 0.76, 1.04) for Flexion; 0.86 (95% CI: 0.70, 1.01) for Extension; 0.87 (95% CI: 0.72, 1.01) for Rotation with Flexion; 0.85 (95% CI: 0.68, 1.02) for Rotation with Extension; 0.68 (95% CI: 0.54, 0.83) for Rotation category.

3.2. Agreement" vs. "disagreement" judgments

We examined the frequency of "agreement" vs. "disagreement" judgments compared to the majority judgments for each classification category (Table 5). Raters were most likely to agree on the classification of Flexion (100%) and least likely to agree on the classification of Rotation (84%). For the Rotation category, when one rater made a judgment of Rotation, the second judgment was Rotation for 84% of the pair-wise comparisons; of the remaining pair-wise comparisons, 4% of the second judgments were Rotation/ Flexion and 12% were Rotation/Extension. For the classifications of Extension, Rotation/Flexion and Rotation/ Extension, the percent agreement among raters was 90.5%, 90.4% and 96.6%, respectively (Table 5).

To explore the nature of the disagreements in the classification decisions, we examined the percentage of disagreements pair-wise raters judgments compared to the majority judgments for each of 15 possible classification disagreements (see Table 6 for all possible combinations) for the 13 raters. The overall percentage of pair-wise classification judgments that disagreed was 12.6% (Table 6). Of the 12.6% of the discordant of pair-wise classification judgments, the most common disagreement occurred with one rater making a judgment of Rotation and the other rater making a judgment of Rotation/Extension (4.2%), followed by the disagreement between the Rotation and Rotation/Flexion categories (2.2%). The least common pair-wise disagreement occurred between judgments of Flexion and Rotation/Extension categories (0%) and between Extension and Rotation/Flexion categories (0%). It is important to note that there was one case for which the expert rater was not in agreement with the majority of classifications.

4. Discussion

This study examined the inter-rater reliability in applying the classification criteria of the MSI approach to clinical data that was previously collected and also explored the pattern of disagreements in classifications that existed in order to determine which categories were

most difficult to distinguish. The reliability for novice users was comparable to the more experienced expert rater and with other studies that have examined the reliability of the MSI approach (Trudelle-Jackson et al., 2008; Harris-Hayes and Van Dillen, 2009). The percent of disagreements in classification judgments, compared to majority ratings, was highest with the Rotation category, involving mismatches with either the Extension/Rotation or Flexion/Rotation categories.

4.1. Inter-rater reliability for MSI approach

The majority of the 13 UVM raters had little to no prior experience with the MSI approach and with training, all raters were able to demonstrate good inter-rater reliability in applying the classification criteria of the this approach with an overall 87.4% agreement in the pairs of classification judgments and a kappa coefficient of 0.81. These results were similar to those reported by others (Trudelle-Jackson et al., 2008; Harris-Hayes and Van Dillen, 2009) and demonstrate that following careful training, the MSI classification system can be learned and thus, has clinical utility.

Other studies have used similar methodology to ours in which paper cases are used rather than actual patients. In their study, Dankaerts et al. (2006) distributed 25 cases (patients' subjective information and videotaped functional tests) to 13 clinicians who independently classified each case using particular pieces of the history and physical examination to classify patients based on a mechanism-based classification system for non-specific LBP (O'Sullivan, 2005). Kappa-coefficients ranged from 0.47 to 0.80 and the percent agreement among raters was 70% (range 60–84%) indicating moderate reliability for this classification approach. These types of studies allow researchers to determine which categories may need further clarification without having to subject patients with LBP to repetitive testing which can aggravate their symptoms, and thus alter subsequent test findings. Furthermore, the use of paper cases removes additional confounding factors such as exam performance by the clinician that can influence reliability outcomes and thus allows the examination of the classification algorithm itself. Other studies, in contrast, have actually had raters examine patients repeatedly (Wilson et al., 1999; Heiss et al., 2004; Petersen et al., 2004; Fritz et al., 2006; Trudelle-Jackson et al., 2008; Harris-Hayes and Van Dillen, 2009; Vibe Fersum et al., 2009). The same patients are examined by different clinicians to determine if similar clinical results are obtained and if assignment to a particular classification category by different examiners is the same for each patient. For the MSI approach, studies have shown percent agreement ranging from 75 to 83% and kappa coefficients from 0.61 to 0.75 when patients are examined by different physical therapists (Trudelle-Jackson et al., 2008; Harris-Hayes and Van Dillen, 2009). While this methodology may simulate actual clinical decision making more closely than written cases, it does not allow direct testing of mutual exclusivity of the subgroups within the algorithm. If the reliability of a classification schema is shown to be poor and there is a high percent of disagreements in classification judgments, one does not know if this is due to a lack of robustness of the algorithm or due to poor reliability of any number of individual test items that are used the algorithm to make the classification determination.

4.2. Patterns of classification disagreements

In our study, there was an overall percentage of discordant pairwise judgments of 12.6% (201 discordant pair-wise raters' judgments), indicating that some patient's with LBP are difficult to categorize with the MSI classification decision-making rules (Harris-Hayes and Van Dillen, 2009). Our results point to two more prevalent patterns of disagreement involving the Rotation category: Rotation – Extension/Rotation and Rotation – Flexion/Rotation mismatches. The Rotation category appears to be the most prevalent MSI subgroup based on a recent clinical trial that examined short term outcomes for patients classified

using the MSI approach (Henry et al., 2012). The percentage of patients classified in the Rotation category was 77%, in Extension/Rotation was 22.5%, and in Flexion/Rotation was 4.6%. Thus, if this prevalence is observed in other studies, it suggests that additional refinement of the Rotation category should be a high priority to minimize misclassifications of patients with LBP. In other studies that have examined mutual exclusivity of classification algorithms as many as 25% of the participants met the criteria for more than one subgroup (Stanton et al., 2011). Thus, the MSI algorithm appears to have a lower prevalence of mismatches at 12.6%.

It is important to note that there was one case for which the expert rater was not in agreement with the majority of classifications. This occurred when the expert classified that patient as “Unclassifiable” and the 13 raters had assigned either Rotation (2 raters), Flexion (1), Flexion/Rotation (8) and Unclassifiable (2). The MSI classification rules (Harris-Hayes and Van Dillen, 2009) are primarily predicated upon the clinician identifying consistent patterns of trunk movements or alignments across the exam that are associated with the patient's low back symptoms. In instances where the patient does not report symptoms during the exam, the clinician must resort to using only visual and tactile judgments about the patient's preferred trunk alignment or movement patterns which are inherently less reliable. In the case where the expert categorized the case as “Unclassifiable,” the patient's report of symptoms was negligible during testing, making this a difficult case to classify.

4.3. Study limitations

While there was an attempt to balance the cases selected across the five MSI categories, with so many categories, there were only three to five cases in each category. Based on the expert rater's judgment as well as the majority rating, the sampling of cases was biased slightly toward the Extension or Rotation/Extension groups (Table 3). Given that percent of agreement is influenced by prevalence of categories, raters had more opportunities to correctly (or incorrectly) classify cases for Extension or Rotation/Extension categories.

The raters' did not examine the patients themselves; instead they made their classification judgments solely on data previously recorded by a physical therapist. While this removed one source of potential error (i.e., inconsistency of clinical exam performance), this did not replicate the decision making that occurs in the clinic. Thus, other factors that impact clinical decision making, which may have improved or worsened the raters' classification success, cannot be taken into account. However, this study design allowed us to examine the reliability of and the pattern of disagreements that existed in the MSI algorithm.

5. Conclusion

The MSI classification algorithm can be learned by novice users and with training, their inter-rater reliability in applying the algorithm for classification judgments is good and similar to that reported in other studies (Trudelle-Jackson et al., 2008; Harris-Hayes and Van Dillen, 2009). However, some degree of error persists in the classification decision-making associated with the MSI system, in particular for the Rotation category.

Additional refinement through the identification of additional and/or novel patient characteristics is needed, especially given that this category appears to be prevalent in the nonspecific LBP population seeking physical therapy treatment.

Acknowledgments

The authors would like to acknowledge the Vermont physical therapy practices that participated in this study: Dee Physical Therapy, South Burlington; Evolution Physical Therapy and Yoga, Burlington; Fletcher Allen Health

Care, Burlington; and Timberlane Physical Therapy, South Burlington. They would also like to acknowledge the assistance of Juvena R. Hitt who compiled and tabulated these data. This research project was funded by the College of Nursing and Health Sciences Dean's Research Incentive Fund, University of Vermont awarded to SMH and by an NIH grant (1R01 HD047709) awarded to LVD.

References

- Dankaerts W, O'Sullivan P. The validity of O'Sullivan's classification system (CS) for a sub-group of NS-CLBP with motor control impairment (MCI): overview of a series of studies and review of the literature. *Manual Therapy*. 2011; 16(1):9–14. [PubMed: 21094624]
- O'Sullivan P. Diagnosis and classification of chronic low back pain disorders: maladaptive movement and motor control impairments as underlying mechanism. *Manual Therapy*. 2005; 10(4):242–55. [PubMed: 16154380]
- Atlas SJ, Deyo RA, Patrick DL, Convery K, Keller RB, Singer DE. The Quebec task force classification for spinal disorders and the severity, treatment, and outcomes of sciatica and lumbar spinal stenosis. *Spine*. 1996; 21(24):2885–92. [PubMed: 9112713]
- Billis EV, McCarthy CJ, Oldham JA. Subclassification of low back pain: a cross-country comparison. *European Spine Journal*. 2007; 16(7):865–79. [PubMed: 17576604]
- Borkan JM, Cherkin DC. An agenda for primary care research on low back pain. *Spine*. 1996; 21(24):2880–4. [PubMed: 9112712]
- Buchbinder R, Goel V, Bombardier C, Hogg-Johnson S. Classification systems of soft tissue disorders of the neck and upper limb: do they satisfy methodological guidelines? *Journal of Clinical Epidemiology*. 1996; 49(2):141–9. [PubMed: 8606315]
- Croft, P. Low back pain.. In: Stevens, A.; Raftery, J., editors. *Health care needs assessment*. Radcliffe Medical Press Oxford; New York, NY: 1997. p. 129
- Dankaerts W, O'Sullivan PB, Straker LM, Burnett AF, Skouen JS. The inter-examiner reliability of a classification method for non-specific chronic low back pain patients with motor control impairment. *Manual Therapy*. 2006; 11(1):28–39. [PubMed: 15936976]
- Fritz JM, Irrgang JJ. A comparison of a modified Oswestry low back pain disability questionnaire and the Quebec back pain disability scale. *Physical Therapy*. 2001; 81(2):776–88. [PubMed: 11175676]
- Fritz JM, Brennan GP, Clifford SN, Hunter SJ, Thackeray A. An examination of the reliability of a classification algorithm for subgrouping patients with low back pain. *Spine*. 2006; 31(1):77–82. [PubMed: 16395181]
- Fritz JM, Cleland JA, Childs JD. Subgrouping patients with low back pain: evolution of a classification approach to physical therapy. *The Journal of Orthopaedic and Sports Physical Therapy*. 2007; 37(6):290–302. [PubMed: 17612355]
- Harris Hayes M, Van Dillen LR, Sahrman SA. Classification, treatment and outcomes of a patient with lumbar extension syndrome. *Physiotherapy Theory and Practice*. 2005; 21(3):181–96. [PubMed: 16389699]
- Harris-Hayes M, Van Dillen LR. The inter-tester reliability of physical therapists classifying low back pain problems based on the movement system impairment classification system. *The Journal of Injury, Function and Rehabilitation*. 2009; 1(2):117–26.
- 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. *The Journal of Orthopaedic and Sports Physical Therapy*. 2004; 34(8):430–9. [PubMed: 15373006]
- Henry SM, Ouellette R, Hitt J, Van Dillen L, Bunn J. Short-term disability and pain outcomes do not differ for patients with low back pain who receive treatment-based classification stabilization or movement system impairment exercises (Abstract). *The Journal of Orthopaedic and Sports Physical Therapy*. 2012; 42(1):A84.
- Jensen MP, Turner JA, Romano JM. What is the maximum number of levels needed in pain intensity measurement? *Pain*. 1994; 58(3):387–92. [PubMed: 7838588]
- Kent P, Keating JL. Classification in nonspecific low back pain: what methods do primary care clinicians currently use? *Spine*. 2005; 30(12):1433–40. [PubMed: 15959374]
- King, JE. *Software solutions for obtaining a kappa-type statistic for use with multipleraters*. Southwest Educational Research Association; Dallas, TX: 2004.

- Koumantakis GA, Watson PJ, Oldham JA. Trunk muscle stabilization training plus general exercise versus general exercise only: randomized controlled trial of patients with recurrent low back pain. *Physical Therapy*. 2005; 85(3):209–25. [PubMed: 15733046]
- Maluf KS, Sahrman SA, Van Dillen LR. Use of a classification system to guide nonsurgical management of a patient with chronic low back pain. *Physical Therapy*. 2000; 80(11):1097–111. [PubMed: 11046197]
- Mannion AF, Mntener M, Taimela S, Dvorak J. A randomized clinical trial of three active therapies for chronic low back pain. *Spine*. 1999; 24(23):2435–48. [PubMed: 10626305]
- McKenzie, R.; May, S. *The lumbar spine: mechanical diagnosis & therapy*. Vol. 1. Spinal Publications New Zealand Ltd; Waikanae, New Zealand: 2003.
- Moffroid MT, Haugh LD, Henry SM, Short B. Distinguishable groups of musculo-skeletal low back pain patients and asymptomatic control subjects based on physical measures of the NIOSH Low Back Atlas. *Spine*. 1994; 19(12):1350–8. [PubMed: 8066515]
- Mooney V. The classification of low back pain. *Annals of Medicine*. 1989; 21(5):321–5.
- Nordin, M.; Weiser, S.; van Doorn, JW.; Hiebert, R.; Rom, WN. *Nonspecific low back pain, environmental and occupational medicine*. Lippincott-Raven Publishers; Philadelphia, PA: 1998. p. 947
- Petersen T, Olsen S, Laslett M, Thorsen H, Manniche C, Ekdahl C, et al. Inter-tester reliability of a new diagnostic classification system for patients with nonspecific low back pain. *Australian Journal of Physiotherapy*. 2004; 50(2):85–94. [PubMed: 15151492]
- Riddle DL. Classification and low back pain: a review of the literature and critical analysis of selected systems. *Physical Therapy*. 1998; 78(7):708–37. [PubMed: 9672545]
- Sahrman, SA. *Movement impairment syndromes of the lumbar spine, diagnosis and treatment of movement impairment syndromes*. Mosby, Inc; St. Louis, MO: 2002. p. 5
- Spitzer WO, LeBlanc FE, Dupuis M. Scientific approach to the assessment and management of activity-related spinal disorders. A monograph for clinicians. Report of the Quebec Task Force on Spinal Disorders. *Spine*. 1987; 12(7):S1–59. [PubMed: 2961086]
- Stanton TR, Fritz JM, Hancock MJ, Latimer J, Maher CG, Wand BM, et al. Evaluation of a treatment-based classification algorithm for low back pain: a cross-sectional study. *Physical Therapy*. 2011; 91(4):469–509.
- Trudelle-Jackson E, Sarvaiya-Shah SA, Wang SS. Interrater reliability of a movement impairment-based classification system for lumbar spine syndromes in patients with chronic low back pain. *Journal of Orthopaedic and Sports Physical Therapy*. 2008; 38(6):371–6. [PubMed: 18515957]
- Van Dillen LR, Sahrman SA, Norton BJ, Caldwell CA, Fleming DA, McDonnell MK, et al. Reliability of physical examination items used for classification of patients with low back pain. *Physical Therapy*. 1998; 78(9):979–88. [PubMed: 9736895]
- Van Dillen LR, Sahrman SA, Norton BJ, Caldwell CA, McDonnell MK, Bloom NJ. The effect of modifying patient-preferred spinal movement and alignment during symptom testing in patients with low back pain: a preliminary report. *Archives of Physical Medicine and Rehabilitation*. 2003a; 84(3):313–22. [PubMed: 12638097]
- Van Dillen LR, Sahrman SA, Norton BJ, Caldwell CA, McDonnell MK, Bloom NJ. Movement system impairment-based categories for low back pain: stage 1 validation. *Journal of Orthopaedic and Sports Physical Therapy*. 2003b; 33(3):126–42. [PubMed: 12683688]
- Van Dillen LR, Sahrman SA, Wagner JM. Classification, intervention, and outcomes for a person with lumbar rotation with flexion syndrome. *Physical Therapy*. 2005; 85(4):336–51. [PubMed: 15794704]
- Van Dillen LR, Maluf KS, Sahrman SA. Further examination of modifying patient-preferred movement and alignment strategies in patients with low back pain during symptomatic tests. *Manual Therapy*. 2009; 14(1):52–60. [PubMed: 18032090]
- Vibe Fersum K, O'Sullivan PB, Kvale A, Skouen JS. Inter-examiner reliability of a classification system for patients with non-specific low back pain. *Manual Therapy*. 2009; 14(5):555–61. [PubMed: 18838331]
- Wilson L, Hall H, McIntosh G, Melles T. Intertester reliability of a low back pain classification system. *Spine*. 1999; 24(3):248–54. [PubMed: 10025019]

Table 1

Characteristics of the 13 raters who evaluated the 21 patient cases.

Number of raters	13 Raters (12 of whom were licensed physical therapists)
Entry level physical therapy education for 12 raters	BS PT = 8; MPT or MSPT = 4
Years in physical therapy practice (mean $\pm$ SD; range)	13 $\pm$ 7.4; 4–27 years
Number of raters with specialty certification from ABPTS	0
% FTE of time in physical therapy practice (n)	
0–25%	4
26–50%	0
51–75%	2
76–100%	6
Previous experience with the MSI Classification system (n)	
None	6
A little	5
Moderate	2
A lot	0

Abbreviations: ABPTS, American Board of Physical Therapy Specialties; MSI, Movement System Impairment.

Table 2

Descriptive characteristics of the 21 patients with LBP who were being classified using the movement system impairment approach.

Characteristic	Mean (SD)
Age (years)	32 (14)
Duration of LBP years)	6.3 (5.4)
Sex (% female)	52
Average LBP symptoms prior 7 days on verbal numeric pain rating scale (0–10) (Jensen et al., 1994)	3.94 (1.55)
Oswestry disability score (0–100%) (Fritz and Irrgang, 2001)	15 (5.28)
Location of symptoms (%)	
Low back	63
Low back and proximal lower extremity	32

Values represent mean (standard deviation) unless otherwise indicated.

Table 3

Distribution of classification judgments for the 13 UVM raters and the number of cases classified by the expert (LVD).

Distribution of classification judgments by 13 raters		# of cases classified by expert (LVD)
Flexion	16% (43)	3
Extension	21.6% (58)	5
Rotation/flexion	18.2% (49)	3
Rotation/extension	26% (70)	5
Rotation	17.5% (47)	4
Unclassifiable	0.7% (2)	1

Total # of classification judgments for 21 cases = 269 judgments

Table 4

The overall percent agreement among the 13 raters and kappa by classification category.

Overall percent agreement	
	Rater agreement (<i>n</i> = 13)
Kappa (95% CI)	0.81 (0.79 - 0.83)
Overall percentage agreement	87.4%
Kappa (95% CI) by classification category	
Flexion	0.90 (0.74, 1.04)
Extension	0.86 (0.70, 1.01)
Rotation/flexion	0.87 (0.72, 1.01)
Rotation/extension	0.85 (0.68, 1.02)
Rotation	0.68 (0.54, 0.83)

Table 5

Distribution of classification judgments and the frequency of “agreement” vs. “disagreement” judgments compared to the majority judgments for each classification category.

Majority classification category ↓	Distribution of classification judgments					
	Flexion	Extension	Rotation/flexion	Rotation/extension	Rotation	Unclassifiable
Flexion	39 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Extension	3 (4.8%)	57 (90.5%)	0 (0%)	1 (1.6%)	2 (3.2%)	0 (0%)
Rotation/flexion	1 (1.9%)	0 (0%)	47 (90.4%)	0 (0%)	2 (3.8%)	2 (3.8%)
Rotation/extension	0 (0%)	1 (1.5%)	0 (0%)	63 (96.6%)	1 (1.5%)	0 (0%)
Rotation	0 (0%)	0 (0%)	2 (4.0%)	6 (12.0%)	42 (84.0%)	0 (0%)
Unclassifiable	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Table 6

Percentage (and number) of discordant pair-wise raters' judgments compared to the majority judgments for each of 15 possible classification disagreements.

Percentage of discordant pair-wise judgments	
Overall percentage of pair-wise judgments that disagreed	12.6% (201)
Percentage of pair-wise judgments that disagreed between:	
Rotation – flexion	0.5% (8)
Rotation – extension	0.9% (14)
Rotation – rotation/extension	4.2% (66)
Rotation – rotation/flexion	2.2% (35)
Rotation – unclassifiable	0.3% (4)
Flexion – extension	1.3% (21)
Flexion – rotation/extension	0% (0)
Flexion – rotation/flexion	0.5% (8)
Flexion – unclassifiable	0.1% (2)
Extension – rotation/extension	1.5% (24)
Extension – rotation/flexion	0% (0)
Extension – unclassifiable	0% (0)
Rotation/extension – rotation/flexion	0.2% (3)
Rotation/extension – unclassifiable	0% (0)
Rotation/flexion – unclassifiable	1% (16)