Published in final edited form as:

Arch Phys Med Rehabil. 2011 June ; 92(6): 913–919. doi:10.1016/j.apmr.2010.12.042.

Low Inter-rater Reliability of Examiners Performing the Prone Instability Test, a Clinical Test for Lumbar Shear Instability

Margaret M. Ravenna, BS, SPT, Shannon L. Hoffman, PT, DPT, and Linda R. Van Dillen, PT, PhD

Musculoskeletal Analysis Laboratory, Program in Physical Therapy, Washington University in St. Louis, School of Medicine

Abstract

Objective—To test the inter-rater reliability of examiners performing the prone instability test (PIT), a clinical test proposed to identify lumbar shear instability.

Design—Cross-sectional, test-retest design examining subjects with mechanical low back pain (LBP).

Setting—University-based musculoskeletal analysis laboratory.

Participants—Thirty subjects with mechanical LBP recruited from community sources in a metropolitan region.

Intervention—Not applicable.

Main Outcome Measures—Repeated measures of a clinical examination test proposed to identify lumbar shear instability.

Results—Inter-rater reliability of examiners' judgments of the PIT results were indexed with percent agreement and the kappa statistic. Examiners obtained 63 % agreement and a kappa value of 0.10 (95% CI: -0.27 to 0.47). Adjusted kappa values based on prevalence and bias indices were calculated to evaluate the effect on kappa. The prevalence index associated with examiner judgments of the PIT was 0.43 and the bias index was 0.03. The prevalence-adjusted-bias-adjusted kappa value was slightly higher than the unadjusted kappa value ($k=0.27$; 95% CI: -0.08 to 0.61).

Conclusions—The results of our study are not consistent with those of previous studies examining reliability of therapists performing the PIT. We conclude that examiners do not attain acceptable inter-rater reliability when performing the procedures for the PIT based on the

© 2011 The American Congress of Rehabilitation Medicine. Published by Elsevier Inc. All rights reserved.

Contact Information: Linda R. Van Dillen Washington University School of Medicine, Program in Physical Therapy Campus Box 8502 4444 Forest Park Blvd St. Louis, MO 63108 Telephone (314) 286 1442 Fax (314) 286 1410 vandillenl@wustl.edu.

Financial Disclosure: We certify that no party having a direct interest in the results of the research supporting this article has or will confer a benefit on us or on any organization with which we are associated, and we certify that all financial and material support for this research and work are clearly identified in the title page of the manuscript.

Device Status Statement: The manuscript submitted does not contain information about medical devices.

This information was presented in poster format at the 2010 Missouri Physical Therapy Association Conference, April 16 – 18, 2010, Clayton, MO and will be presented in poster format at the Combined Sections Meeting of the American Physical Therapy Association, February 9–12, 2011, New Orleans, LA.

Reprints will not be available from the authors

Clinical trial registration number: NCT00802724

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.

information that is currently provided in the literature. Based on our experience we suggest further exploration, standardization, and clarification of procedural details to improve therapists' ability to conduct the PIT on people with LBP.

Keywords

Diagnosis; Low back pain; Rehabilitation; Reliability

A common complaint of people who experience mechanical low back pain (LBP) is an increase in symptoms when bending forward. These symptoms are believed to be the result of stress and injury of lumbar region tissue due to repeated lumbar flexion. Recent biomechanical data, however, suggests that for some individuals the increase in symptoms may be the result of increased anterior shear forces on one or more lumbar vertebrae.¹ McGill has reported that the typical line of action of the lumbar extensor muscles is posterior and caudal, producing a posterior shear force that resists anterior shear of the lumbar vertebrae during trunk flexion. McGill et al has proposed that during trunk flexion, the oblique line of action of the lumbar extensors realigns to the compressive axis of the spine.^{1,2} The proposed result is a loss of resistance to anterior shear forces in the lumbar region and, over time, potential tissue damage.²

The prone instability test (PIT) has been described by several individuals. It was originally described to identify people who demonstrate lumbar segmental instability and was called the segmental instability test.^{3,4} McGill has described this test as a manual test for lumbar joint shear stability.¹ Recently, authors have referred to this test as the PIT.⁵⁻⁷ The test requires the person to lie prone on an examination table with the trunk supported and relaxed, hips flexed over the edge of the table, and the feet positioned on the floor. In this position (Position 1) an examiner applies a passive posterior-anterior (PA) force over each lumbar spinous process. Symptoms with each PA force are obtained. If a person reports an increase in symptoms with a PA force, he is instructed to hold onto the table with his arms and lift his legs to activate the lumbar extensor muscles (Position 2). A PA force is reapplied to the symptomatic lumbar vertebra and symptoms are compared to those with the first PA force. If symptoms are decreased with the second PA force, the test is positive for the contribution of lumbar vertebral shearing at that level to his symptoms. A positive finding at one or more vertebral levels yields an overall positive result for the PIT.

The theoretical basis for the PIT is that with the paraspinals relaxed in Position 1, the examiner can produce an anterior shear force to each lumbar vertebra and determine a person's symptom response to the shear force. In Position 2, the activation of the lumbar paraspinals should prevent anterior shearing of the lumbar vertebrae. If anterior shearing was the cause of the person's symptoms in Position 1, their symptoms should then be alleviated in Position 2.¹

Three studies have documented the inter-rater reliability of examiners performing the PIT in people with LBP. In two studies, reliability was assessed for examiners' judgments of whether or not people displayed an overall positive PIT. In both studies, moderate to high agreement⁸ was obtained ($k=.69^5$ and $.87^6$ respectively). In a third study, reliability of the PIT was examined in two stages.⁷ The first stage of testing assessed the reliability of examiners performing the procedures in Position 1 (i.e., application of the PA force and interpretation of symptoms at each vertebral level). The second stage of testing assessed the reliability of examiners performing the same procedures in Position 2 for each vertebral level that was symptomatic in Position 1. Moderate agreement⁸ was obtained for testing in Position 1 ($k=.54$) and Position 2 ($k=.46$).

Our original interest was to understand the relationship between lumbar shear instability and findings from other aspects of a clinical examination. We also wanted to identify characteristics of subjects who displayed positive findings for lumbar shear instability. Because of the variable findings in prior reliability studies and because some studies⁵⁻⁷ involved people from the same research group, we considered it important to test the reliability of our novice examiners before using the PIT in subsequent studies. Agreement on standardized procedures for performance and interpretation of findings with the PIT has important clinical implications. Findings from recent studies suggest a positive PIT is the most important test for classifying people who will benefit from a stabilization exercise program.⁹ Therefore, it is important to ensure that the test can be replicated consistently and interpreted appropriately so future examiners can accurately direct treatment for individuals with LBP.

The purpose of the current study was to examine the inter-rater reliability of novice examiners performing the PIT. We hypothesized that clinically acceptable reliability¹⁰ would be attained when the test was administered by trained examiners.

METHODS

Participants

Thirty people with chronic or recurrent LBP¹¹ were recruited for reliability testing of the PIT from community sources in a metropolitan region (Table 1). Subjects were between the ages of 18 and 60 years and were required to have current complaints of their typical LBP, but not of an acute flare-up.¹¹ Exclusion criteria included body mass index greater than 30 kg/m², disc herniation, radiating symptoms below the knee, lower extremity weakness or loss of reflexes, history of spinal surgery or fracture, spinal deformity, systemic inflammatory condition, neurological disease, or other serious medical conditions. People with LBP attributable to pregnancy or a primary hip problem were also excluded. Forty-one additional subjects with or without LBP were recruited prior to reliability testing for the training phases of the study (see *Training of Examiners* below). The testing protocol was approved by the university's Human Research Protection Office and all subjects provided written informed consent prior to participation.

Training of Examiners

Two examiners, a second-year Doctor of Physical Therapy student and a licensed physical therapist with two years of clinical experience in outpatient orthopedic physical therapy, participated in this study. Training of the examiners occurred in two phases. During the first phase, twenty subjects with or without LBP were piloted to investigate procedures of the PIT and to provide training in basic skills required to administer the test. Current literature^{1,3-7} and a previous author (Gregory E. Hicks PT, PhD, 6/10/09, 6/23/09, 8/17/09) were consulted, decision-making rules were clarified or established, and a manual of operations was assembled to standardize procedures. During the second phase of training, twenty-one people with LBP were tested. Examiners first tested subjects together and then separately, discussing each case to ensure consistent performance of procedures and interpretation of the test.

Additional Guidelines for Performing and Interpreting the PIT

A basic assumption of the PIT is that the paraspinals are relaxed in Position 1 so that sufficient shearing of the vertebrae can occur and a symptom response can be obtained.¹ However, when following the basic procedures of the PIT as outlined thus far in the literature, we found that some subjects were unable to relax their paraspinals in Position 1 (verified by palpation). The current literature did not indicate how to proceed with the test in

this case. Therefore, we made certain standardized modifications to the procedures, such as adding a stabilization belt and allowing shorter subjects to rest their feet on a foot stool if their feet were unable to reach the ground comfortably. Additionally, there was conflicting information in the literature regarding whether or not subjects flex their knees during the performance of the test.^{1,3,5,6,7} Figure 1 depicts our test set-up for both Position 1 and Position 2. Detailed procedures for performance of the PIT for this study and rationale for any standardization of procedures that were ambiguous as defined in the current literature were documented and are outlined in the Appendix.

We also encountered two significant questions concerning the interpretation of the findings of the PIT that had not been described by previous authors. First, subjects' ability to discriminate between a provocation of LBP and a painful superficial pressure with the first PA in Position 1 was identified as an important factor influencing the interpretation of the PIT. Sometimes subjects reported a painful superficial pressure related to the contact between the examiner's hand and the subject's vertebra, described as "bone-on-bone" pressure. At other times, subjects reported LBP deep to, and significantly different from, this "bone-on-bone" pressure. We decided that the painful superficial pressure was not a true provocation of LBP. Position 2 of the PIT was only performed if the subject reported that LBP was produced with the application of the first PA in Position 1. If the subject reported only painful superficial pressure with the first PA in Position 1, Position 2 was not performed, and a negative result for the PIT at that vertebral level was recorded.

Second, it was unclear how to interpret the findings if simply raising the lower extremities to assume Position 2 increased the subject's LBP symptoms. We decided that if a subject reported a worsening of LBP when lifting his feet no more than 2 inches off of the ground, the PIT result was negative. The primary rationale for this decision was that Position 2 of the PIT should theoretically relieve, not cause, the LBP symptoms of a person with lumbar shear instability. However, because of the lack of clarity in the literature regarding this issue, the second PA was still applied and subjects were asked to assess their symptoms relative to the first PA, regardless of their symptoms with leg lifting. These procedures allowed us to analyze and interpret the results with and without regard to reproduction of LBP symptoms with leg lifting.

Details regarding the interpretation of the PIT, including rationale, for this study can also be found in the Appendix.

Procedures

Examiner order was randomized. The first examiner obtained a general LBP history and performed the PIT. The subject then completed questionnaires concerning his current and past LBP,¹²⁻¹⁶ physical activity level,¹⁷ and general health.¹⁸ Completion of questionnaires required approximately 20 minutes. The second examiner, blinded to the results of the first examiner, then administered the second PIT.

Data Analysis

Descriptive statistics were calculated on variables related to subject characteristics. Inter-rater reliability of the overall judgment for the PIT was examined under two conditions: (1) when the examiners used the rule pertaining to symptoms with leg lifting, and (2) when examiners disregarded this rule (see *Additional Guidelines for Interpreting the PIT*). Note that a positive finding at one or more vertebral levels yielded an overall positive result for the PIT. Percent agreement and the kappa statistic were used to index inter-rater reliability. Kappa was used to evaluate the agreement between the two examiners and corrected for the amount of agreement that was expected to occur by chance alone.¹⁹ If there was an unequal

distribution of responses across categories or if raters were biased to a particular outcome, however, kappa could have been affected.²⁰ Therefore, prevalence and bias indices were also calculated.²¹ Finally, a prevalence-adjusted-bias-adjusted kappa (PABAK) coefficient was calculated to provide a statistical adjustment of kappa given the prevalence and bias associated with testing.²¹ Kappa and PABAK values indexing reliability were interpreted as follows: less than 0.0 is poor; 0.0 to .20 is slight; .21 to .40 is fair; .41 to .60 is moderate; .61 to .80 is substantial; and .81 to 1.0 is almost perfect.⁸

RESULTS

Examiners obtained 63% agreement and a slight⁸ kappa value when performing the PIT (Table 2). The distribution of examiner judgments for the overall PIT results when applying the leg lifting rule is presented in Table 3. Approximately 70% of PIT findings were positive for each examiner (Table 3). The prevalence index was high and the bias index was low. The PABAK value, taking prevalence and bias into account, was slightly higher than the unadjusted kappa value and interpreted as fair.⁸

When the leg lifting rule was disregarded, more subjects had a positive overall PIT result (83% for each examiner) (Table 4). This analysis resulted in 73% agreement and a slight⁸ kappa value (Table 2). The prevalence index was high and the bias index was low. The value of the PABAK statistic was higher than the unadjusted kappa and was interpreted as moderate.⁸

DISCUSSION

The purpose of the current study was to test the inter-rater reliability of examiners performing the PIT. Contrary to previous studies, our examiners were not involved in the early studies⁵⁻⁷ examining the reliability of examiners to perform the PIT. Our raters demonstrated low inter-rater reliability when required to make judgments about findings from the PIT. We believe our experience is representative of novice clinicians learning to perform and interpret the PIT. Studies of other clinical tests intended for use in diagnosing and directing treatment of people with low back pain have also documented low inter-rater reliability and many factors have been found to contribute to the variability in judgments.²² Factors contributing to our results include the effect of prevalence on kappa, addressing details of the PIT not previously documented, feasibility of the test protocol, and differences in sample characteristics compared to previous studies.

We believe our low kappa values were partially due to a skewed distribution of subjects' responses. The prevalence indices were high. Examination of the data indicated that the proportion of positive findings was substantially greater than that of negative findings for each examiner, regardless of whether or not the leg lifting rule was applied. The bias indices were low indicating that systematic bias of an examiner did not influence the results. High prevalence often results in lower kappa values, regardless of percent agreement values.²⁰ To examine the effects of prevalence on kappa values, PABAK values were calculated, and these values were higher than the original kappa (Table 2). PABAK calculations were utilized in a previous PIT reliability study⁷ and have been documented as a common way to statistically compare the effects of prevalence and bias on kappa.²¹ However, PABAK values in this study were still only fair or moderate⁸. Additionally, PABAK values should be considered with caution because some experts consider these values to be inflated and not the best measurement of reliability.²³ Despite efforts made to use statistical measures to illustrate the effect of prevalence on kappa, it still appears we were unable to reliably discriminate between positive and negative findings for the PIT. Therefore, we believe other

factors, in addition to a high prevalence of positive findings, contributed to our low inter-rater reliability.

The greatest source of variability between examiners was related to symptoms with leg lifting (5 of the 11 cases; 45%). These five subjects reported pain with leg lifting with the first test session, but not the second session. Because the examiner order was randomized the difference in symptoms with leg lifting between test sessions could not be attributed to examiner bias. When the data was analyzed disregarding the rule pertaining to symptoms with leg lifting, examiners demonstrated higher percent agreement, but a lower kappa value compared to when the data was analyzed abiding by the rule. However, more subjects had a positive overall PIT result and the PABAK statistic was increased. Although inter-rater reliability based on percent agreement and the PABAK statistic appeared to improve when the rule was disregarded, the question of how to correctly interpret the PIT when subjects report LBP symptoms with lifting their legs remains. Furthermore, if it is determined that LBP symptoms with leg lifting should be interpreted as a negative finding for the PIT (i.e., the rule is applied), a more effective method for standardizing the height of each leg lift may be warranted. Although subjects were instructed to only lift their legs approximately 2 inches from the ground, this was difficult to control. Variability in report of pain with leg lifting may have been due to repeated lifting if subjects lifted their legs too high on their first attempt to achieve Position 2.

Subjects' discrimination between LBP and a painful superficial pressure with the first PA force could have been an important factor influencing the administration, interpretation, and reliability of examiners performing the PIT. One case showed examiner disagreement in the overall PIT result because of differences in symptom report between examiners (LBP versus painful superficial pressure). However, the overall PIT result was positive if just one vertebral level had a positive finding. At each individual vertebral level (where responses at L1-L5 could each be positive or negative), 9 cases showed examiner disagreements in positive or negative findings due to differences in symptom report. For example, both examiners could have reported a positive result at L5. However, for one examiner, the subject could have reported LBP with the PA in Position 1 at L4, which was decreased with the PA in Position 2, indicating a positive result at L4. For the other examiner, the same subject could have only reported a painful superficial pressure at L4 with the PA in Position 1, indicating a negative result at L4. Therefore, both examiners could have reported an overall positive result because both recorded a positive finding at L5, but they still disagreed on the finding at the L4 Level. Because symptom report in Position 1 dictates the performance and interpretation of the remainder of the test, disagreement at the vertebral level could influence other clinicians' reliability. Clarification of procedures related to the nature of the symptom report (pain versus pressure) in Position 1 is necessary.

The remaining cases of disagreement between examiners can be attributed to procedural and feasibility issues. Differences in (1) positioning of the spine and application of PA forces in subjects with a thoracic kyphosis or large abdomen, (2) the ability of subjects to relax their paraspinal muscles in Position 1, and (3) levels of subject tolerance to the PA forces were potential sources of variability between examiners.

A comparison of subject characteristics between the current study and those previously reported^{5,6} may also explain differences in inter-rater reliability reports (Table 5). Age, proportion of subjects with a history of prior LBP, and Fear Avoidance Beliefs Questionnaire subscale scores were similar between prior studies and the current study. However, compared to prior study samples, the current sample had a higher percentage of men and reported higher levels of disability on the modified Oswestry Disability Questionnaire. Although the impact of gender and disability on the outcome of the PIT is

unknown, these characteristics may provide a possible explanation for why a larger proportion of our subjects had positive PIT results compared to other studies. Subject characteristic data was not available for comparison of subjects in the current study to Schneider et al.⁷

Finally, we acknowledge that our addition of standardized procedures not previously described in the literature to help achieve paraspinal muscle relaxation in Position 1, such as adding a foot stool or stabilization belt, may also explain our low reliability compared to previous studies.⁵⁻⁷ Additionally, we allowed our subjects to flex their knees in Position 1 if necessary to allow for relaxation of their paraspinal muscles. Two of the other reliability studies illustrated the test as performed with the knees fully extended.^{6,7} This difference in procedure could also explain the poor reliability demonstrated in this study. It could be argued that, based on these modifications or differences in procedures, we were not conducting the same test as the other reliability studies. On the other hand, we felt these modifications were necessary to ensure that basic assumptions of the test were being met.¹ The current literature offered no guidance of how to proceed with testing if these assumptions were not met. We feel that this further highlights the need for clarification and standardization of procedures for the PIT in the literature. This is important if a wide range of clinicians are learning to perform this test to assist with diagnosis and treatment.⁹

One limitation of our study is that we excluded subjects with a body mass index greater than 30 kg/m², which may reduce the generalizability of our results. Additionally, the level of clinical experience of our examiners and their familiarity with the PIT was less than examiners in the previous reliability studies.⁵⁻⁷ However, we believe our results are indicative of how novice clinicians using the PIT might perform.

CONCLUSION

Examiners in the current study did not attain clinically acceptable inter-rater reliability when performing the procedures for the PIT. Differences in procedural interpretation and sample characteristics may account for the differences in results between the current study and prior studies. Based on the current study, further standardization and clarification of procedures is warranted to improve therapists' ability to consistently conduct the PIT on people with LBP.

Acknowledgments

We thank Gregory E. Hicks, PhD, PT for his scientific advice regarding the procedures and interpretation of the Prone Instability Test. We thank Kathryn Baxter, research patient coordinator, for her assistance with subject recruitment, screening, and organization of data. We thank Mollie Schlarman, SPT, for her assistance in data entry. We thank Ashley Hay, SPT, for giving us permission to use her photo as a figure in this manuscript and whose photo was also used to illustrate the proper positioning for the PIT when instructing subjects during testing.

Written permission from all persons named in acknowledgement has been obtained.

Grant and Financial Support: National Institutes of Health Grants UL1 RR024992, TL1 RR024995, R01 HD047709-04 and funds from the Missouri Physical Therapy Association.

APPENDIX: Procedures for administering the Prone Instability Test (PIT)

1. The subject lies prone with his torso on an examination table and his legs off the end of the table (Position 1).
 - a. The edge of the table is at or near the inguinal crease of the subject.
 - b. The subject's torso and pelvis should be resting flat on the table, with the lumbar spine parallel to the table (not flexed or extended).

- i. *Rationale:* The subject is positioned in this manner to ensure relaxation of the paraspinals and standardize the position of the lumbar spine.
 - c. The subject's head can be turned to either side.
 - d. The plantar aspects of the metatarsal heads are resting on the floor. A stool may be placed under the feet of the subject if his feet do not comfortably reach the floor. Not all subjects may require a foot stool.
 - i. *Rationale:* The height of the table is not adjustable. Therefore, the foot stool is used when necessary to compensate for a subject's height, to enable all subjects to have a firm foot hold, and to help achieve proper positioning of the lumbar spine, all in an effort to facilitate paraspinal relaxation.
 - e. The subject may bear weight through his legs, as long as his/her paraspinals are relaxed.
 - i. *Rationale:* Some subjects are able to demonstrate better relaxation of their lumbar and thoracic paraspinals if they are able to bear some weight through their legs.
 - f. The subject may be allowed to bend his knees if it allows him to relax his paraspinals more readily and does not alter the position of the lumbar spine (i.e., cause it to flex or extend).
 - i. *Rationale:* With their knees straight, some subjects either go into lumbar extension or demonstrate activation of their paraspinal muscles.
 - g. A trunk stabilizing belt is placed around the subject and table at shoulder level.
 - i. *Rationale:* The trunk stabilizing belt is used to encourage subjects to relax throughout their torsos. Without the stabilizing belt, some subjects report that it feels like they are going to fall off the table and have difficulty fully relaxing their paraspinals.
2. The examiner palpates and labels, with a small sticker, each lumbar spinous process. The stickers are to be removed at the end of testing, prior to completion of the questionnaires, by Examiner 1 to ensure that their placement is not evident to Examiner 2.
 3. The examiner palpates the subject's lumbar and thoracic paraspinals to ensure that they are fully relaxed.
 - a. The subject is repositioned as necessary, within the above guidelines, to achieve paraspinal relaxation.
 - i. *Rationale:* A basic assumption of the test is that the paraspinals are relaxed when the posterior-anterior force (PA) is applied.
 4. The examiner applies a PA, using a steady force similar in force to a Grade IV mobilization,²⁴ at each lumbar vertebra, beginning with L5. The examiner assesses the subject's pain in response to the PA.

- a. If the subject reports no pain with the PA, the test is negative for that vertebral level and the examiner continues testing at the next proximal vertebra.
- b. If the subject reports pain with the PA at a particular vertebral level:
 - i. The subject is asked to determine if the PA produces low back pain symptoms or if it is a superficial bone-on-bone pressure related to pressure on a bony prominence. If the subject reports low back pain symptoms with the PA, the test continues for that vertebral level as described in 4.b.ii. If the subject reports superficial bone-on-bone pressure, the test is negative for that vertebral level and the examiner continues testing at the next proximal vertebra.
 1. *Rationale:* The superficial pressure is not considered a replication of low back pain (LBP) symptoms caused by shearing of the vertebra, but rather a bone-on-bone pressure related to a force exerted on a bony prominence.
 - ii. If the subject reports reproduction of low back pain symptoms at a given vertebral level, the subject is instructed to hold onto the table with his arms and lift his legs no more than 2 inches off of the floor (Position 2). The examiner palpates the lumbar paraspinals to ensure they are activated.
 1. *Rationale:* Subjects are instructed to lift their legs only 2 inches off the floor to ensure that their paraspinals are recruited, but that they are not extending their lumbar spine.
 - iii. The hand of the examiner should remain on the affected vertebral segment as the subject lifts his legs.
 1. *Rationale:* This procedure is followed to ensure that the same vertebra is being tested when the subject lifts his legs.
 - iv. If at any given vertebral level, a subject reports an increase in symptoms with lifting his legs for Position 2, the examiner must ensure that the subject is holding onto the table and lifting his legs no more than 2 inches off the floor. If the subject's performance does not meet these criteria, the subject is asked to lower his legs back to the floor. The subject is then re-instructed to assume Position 2 while holding onto the table and only lifting his legs no more than 2 inches off the floor.
 1. *Rationale:* These procedures are used to ensure that the increase in symptoms is not due to other factors, such as increased lumbar extension.
 - v. If the subject still reports an increase in symptoms with lifting his legs, the result for that vertebral level is

automatically negative. However, a second PA is still applied and subjects are asked to assess their symptoms relative to the first PA, regardless of their symptoms with leg lifting, as described in 4.b.vii.

1. ***Rationale:*** With pilot testing, some subjects complained of increased symptoms when lifting their legs. Due to a lack of clarity in the literature regarding this issue, this rule (the leg lifting rule) has been developed. The primary rationale for this decision is that Position 2 of the PIT should theoretically relieve, not cause, the LBP symptoms of a person with lumbar anterior shear. Still applying the second PA will allow for analysis and interpretation of the results with and without regard to the subjects' reproduction of LBP symptoms with leg lifting.
- vi. If the subject reports no change in symptoms with leg lifting, the test proceeds as described in 4.b.vii.
- vii. The examiner then applies a second PA to the same vertebral segment and asks the subject to compare symptoms with the second PA to the first PA (i.e., increased, decreased, or stayed the same).
- viii. If the subject reports decreased symptoms with the second PA compared to the first PA, the test is positive for that vertebral level. If the subject reports that his symptoms stay the same or increase with the second PA compared to the first PA, the test is negative for that vertebral level.
5. Step 4 is repeated for each lumbar vertebra working proximally from L5 to L1.
6. A positive finding at one or more vertebral levels yields an overall positive result for the PIT.

LIST OF ABBREVIATIONS

LBP	low back pain
PIT	prone instability test
PA	posterior-anterior

REFERENCES

1. McGill, S. Low back disorders: Evidence-based prevention and rehabilitation. 1st ed. Human Kinetics Publishers; Champaign: 2002.
2. McGill S, Hughson RL, Parks K. Changes in lumbar lordosis modify the role of the extensor muscles. Clin Biomech. 2000; 15:777–80.
3. Magee, DJ. Orthopedic Physical Assessment. 4th ed. Elsevier Sciences; Philadelphia: 2006.
4. Wadsworth, CT. Manual Examination and Treatment of the Spine and Extremities. Williams & Wilkens; Baltimore: 1988.

5. Fritz JM, Piva SR, Childs JD. Accuracy of the clinical examination to predict radiographic instability of the lumbar spine. *Eur Spine J*. 2005; 14:743–50. [PubMed: 16047211]
6. Hicks GE, Fritz JM, Delitto A, Mishock J. Interrater reliability of clinical examination measures for identification of lumbar segmental instability. *Arch Phys Med Rehabil*. 2003; 84:1858–64. [PubMed: 14669195]
7. Schneider M, Erhard R, Brach J, Tellin W, Imbarlina F, Delitto A. Spinal palpation for lumbar segmental mobility and pain provocation: an interexaminer reliability study. *J Manipulative Physiol Ther*. 2008; 31:465–73. [PubMed: 18722203]
8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–74. [PubMed: 843571]
9. Hicks GE, Fritz JM, Delitto A, McGill SM. Preliminary development of a clinical prediction rule for determining which patients with low back pain will respond to a stabilization exercise program. *Arch Phys Med Rehabil*. 2005; 86:1753–62. [PubMed: 16181938]
10. Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic*. 1981; 86:127–37. [PubMed: 7315877]
11. Von Korff M. Studying the natural history of back pain. *Spine*. 1994; 19:2041S–6S. [PubMed: 7801181]
12. Deyo RA, Andersson G, Bombardier C, et al. Outcome measures for studying patients with low back pain. *Spine*. 1994; 19:2032S–6S. [PubMed: 7801179]
13. Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain*. 1993; 52:157–68. [PubMed: 8455963]
14. Turpin RS, Ozminkowski RJ, Sharda CE, Collins JJ, Berger ML, Billotti GM, Baase CM, Olson MJ, Nicholson S. Reliability and validity of the Stanford Presenteeism Scale. *J Occup Environ Med*. 2004; 46:1123–33. [PubMed: 15534499]
15. Fritz JM, Irrgang JJ. A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. *Phys Ther*. 2001; 81:776–88. [PubMed: 11175676]
16. Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine*. 1983; 8:141–4. [PubMed: 6222486]
17. Baecke JA, Burema J, Frijters JE. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *Am J Clin Nutr*. 1982; 36:936–42. [PubMed: 7137077]
18. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992; 30:473–83. [PubMed: 1593914]
19. Cohen J. Coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960; 20:37–46.
20. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol*. 1990; 43:543–9. [PubMed: 2348207]
21. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol*. 1993; 46:423–9. [PubMed: 8501467]
22. Hunt DG, Zuberbier OA, Kozlowski AJ, Robinson J, Schultz IZ, Milner RA, Crook JM, Turk DC. Reliability of the lumbar flexion, lumbar extension, and passive straight leg raise test in normal populations embedded within a complete physical examination. *Spine*. 2001; 26:2714–8. [PubMed: 11740361]
23. Hoehler FK. Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *J Clin Epidemiol*. 2000; 53:499–503. [PubMed: 10812322]
24. Maitland, GD. *Vertebral Manipulation*. 5th ed. Butterworth-Heinemann; Oxford: 1986.

Figure 1. Positions 1* and 2[†] of the Prone Instability Test, respectively

*Position1, subject lies relaxed in prone on an examination table with the trunk supported, hips flexed over the edge of the table and the feet positioned on the floor or stool. The first passive posterior-anterior pressure (PA) is applied over each lumbar vertebrae; [†]Position 2, subject is instructed to lift legs 2 inches off the ground to activate the lumbar paraspinal muscles while holding onto table. The second PA is repeated over each lumbar vertebra at which low back pain symptoms were produced in Position 1.

Table 1

Subject Characteristics (N = 30)

Variable	Mean (SD)
Body Mass Index (kg/m ²)	24.5 (3.3)
Age (years)	36.1 (11.8)
Diagnosis (%)	
Degenerative Disc Disease	16.6 (NA)
Disc Problem	10.0 (NA)
LBP*	73.4 (NA)
Report of Previous LBP episodes [†]	83 (NA)
Gender (% male)	56.7 (NA)
Current Pain Intensity (0-10)	2.8 (1.6)
Modified Oswestry [‡] (0-100%)	23.9 (10.0)
FABQ [§] Physical subscale (0-24)	14.3 (3.7)
FABQ [§] Work subscale (0-42)	15.4 (6.5)

* LBP: low back pain

[†] Episode is a period of low back pain that has limited the performance of daily activities for 3 days or greater

[‡] Modified Oswestry Low Back Pain Disability Questionnaire¹⁵

[§] Fear Avoidance Beliefs Questionnaire¹³

Reliability Coefficients for the Prone Instability Test (PIT) Result under Two Conditions Pertaining to Symptoms with Leg Lifting

Table 2

Condition	Percent Agreement	Kappa (95% CI)	PABAK* (95% CI)	Bias Index	Prevalence Index
Acknowledged [†]	63%	0.10 (-0.27,0.47)	0.27 (-0.08,0.61)	0.03	0.43
Ignored [‡]	73%	0.04 (-0.34,0.42)	0.47 (0.15,0.78)	0.00	0.67

* PABAK: Prevalence-Adjusted Bias-Adjusted Kappa

[†] Acknowledged: examiners abided by the rule that if a subject experienced an increase in symptoms lifting his legs into Position 2, the PIT result was automatically coded as negative

[‡] Ignored: examiners ignored the rule that automatically coded the PIT result as negative if the subject experienced an increase in symptoms when lifting his legs into Position 2

Table 3

Distribution of Examiner Judgments for the Prone Instability Test Acknowledging the Rule Pertaining to Symptoms with Leg Lifting*

		Examiner 2 Negative	Examiner 2 Positive	Total
Examiner 1	Negative	3	5	8
Examiner 1	Positive	6	16	22
	Total	9	21	30

* Rule: if a subject reported an increase in symptoms when lifting his legs into Position 2, the PIT result was negative.

Table 4

Distribution of Examiner Judgments for the Prone Instability Test Ignoring the Rule Pertaining to Symptoms with Leg Lifting*

		Examiner 2 Negative	Examiner 2 Positive	Total
Examiner 1	Negative	1	3	4
Examiner 1	Positive	5	21	26
	Total	6	24	30

* Rule: if a subject reported an increase in symptoms when lifting his legs into Position 2, the PIT result was negative.

Table 5

Comparison of Sample Characteristics to Previous Studies Testing the Inter-rater Reliability of the Prone Instability Test

Variable	Ravenna et al N=30	Fritz et al ⁵ N=49	Hicks et al ⁶ N=63
Number of examiners	2	2	4; 3 pairs
Age (years)	36.1 (11.8)*	39.2 (11.3)	36.0 (10.3)
Gender (% male)	56.7	42.9	39.7
Modified Oswestry [†] (0-100%)	23.9 (10.0)	20.4 (13.3)	17.8 (11.3)
Prior LBP [‡] episodes (% yes)	83	84	81
FABQ [§] – Physical subscale (0-24)	14.3 (3.7)	15.0 (6.5)	Not available
FABQ [§] – Work subscale (0-42)	15.4 (6.5)	12.9 (11.8)	Not available
Kappa for overall PIT	0.10	0.69	0.87

* Mean (SD)

[†] Modified Oswestry Low Back Pain Disability Questionnaire¹⁵

[‡] LBP: low back pain

[§] Fear Avoidance Beliefs Questionnaire¹³