Washington University School of Medicine
Digital Commons@Becker

Open Access Publications

2010

Development and preliminary reliability testing of an assessment of patient independence in performing a treatment program: Standardized scenarios

Marcie Harris-Hayes
Washington University School of Medicine in St. Louis

Gregory W. Holtzman
Washington University School of Medicine in St. Louis

Jeanne A. Earley
Rehabilitation Institute of St Louis

Linda R. Van Dillen
Washington University School of Medicine in St. Louis

Follow this and additional works at: http://digitalcommons.wustl.edu/open_access_pubs

ORIGINAL REPORT

DEVELOPMENT AND PRELIMINARY RELIABILITY TESTING OF AN ASSESSMENT OF PATIENT INDEPENDENCE IN PERFORMING A TREATMENT PROGRAM: STANDARDIZED SCENARIOS*

Marcie Harris-Hayes, PT, DPT, OCS[1], Gregory W. Holtzman, PT, DPT[1], Jeanne A. Earley, PT, MHS[2] and Linda R. Van Dillen, PT, PhD[1]

From the [1]Program in Physical Therapy, Washington University Medical School and [2]The Rehabilitation Institute of St Louis, St Louis MO, USA

Background: Physical therapists often assess patient independence through observation; however, it is not known if therapists make these judgments reliably. We have developed a standardized method to assess a patient's ability to perform his or her treatment program independently.
Objectives: To develop a standardized assessment of patient independence in performance of a treatment program and examine the intra- and inter-rater reliability decisions made by two physical therapists.
Design: Test-retest.
Methods: An assessment of patient independence in performance was developed. Standardized patient scenarios were used to assess the intra- and inter-tester reliability of two physical therapists. Percentage of agreement (%) and kappa's coefficient (k and k_w) indexed rater reliability.
Results: Intra-rater reliability of therapist 1 was as follows: knowledge: % = 95, k = 0.90; performance: % = 95, k_w = 0.82. Intra-rater reliability of therapist 2 was as follows: knowledge: % = 85, k = 0.68; performance: % = 94, k_w = 0.80. Inter-rater reliability for knowledge was % = 91 and k = 0.79 and for performance was % = 91 and k_w = 0.72.
Conclusion: Trained therapists displayed substantial to excellent intra-rater reliability and substantial inter-rater reliability in assessing a patient's independence in a treatment program.

Key words: activities of daily living; exercise therapy; directly observed therapy; patient compliance.

INTRODUCTION

Physical therapists (PTs) commonly prescribe specific treatments for their patients with the goal of improving patient outcomes (1). Assuming the treatment is appropriate, improved outcomes are expected if the patient is adherent to the prescription. One proposed prerequisite to patient adherence is the patient's ability to perform the treatment independently. A patient is independent in performance if he or she performs the treatment correctly without any assistance. A patient may perform his or her treatment at the instructed duration, frequency and intensity level; however, if the patient is not independent in the performance of the treatment program as defined, outcomes may be negatively affected. We believe it is as important to assess the patient's ability to perform the treatment independently as it is to assess how often he or she performs the treatment.

In studies of the relationship between treatment and outcomes, parameters such as the duration, frequency and intensity of the treatment are commonly measured (2–7). To measure adherence, self-report questionnaires are typically used to determine the frequency of performance; however, these questionnaires provide no measurement of the quality of performance. Quality of performance is important because the patient may report that they are performing the treatment as prescribed; however, the performance may be suboptimal, i.e., incorrect. In this situation, the patient would not be able to adhere to the prescribed treatment, and treatment would need to be adjusted to the patient's ability level.

The patient's ability to perform treatment independently is often assessed by a physical therapists (PT) through observation. No standardized method to assess performance has been described, particularly for patients with musculoskeletal pain conditions. Standardized methods to assess patient performance may be helpful in decisions regarding treatment prescription and progression, which will improve the treatment effectiveness, and thus, patient outcomes.

To develop standardized methods of assessment, factors that influence the patient's ability to perform the treatment independently must be considered. We propose 2 factors that influence independent performance of treatment; cognition and psychomotor skill. In the current study, cognition refers to the ability of the patient to understand the key concepts underlying the prescribed treatment and how the key concept relates to his or her overall limitations. The key concept refers to the primary goal underlying the exercise or activity of daily living (ADL)

---

prescribed. Psychomotor skill refers to the patient's ability to physically perform the exercise or ADL. Deficits in cognition, psychomotor skill or both could affect independence in performance. Different strategies would be required to address deficits in either of these domains. A standardized assessment to identify the primary factor(s) (cognition or psychomotor skill) contributing to suboptimal performance would be useful and would provide an objective method for determining the best strategy to modify the patient's performance.

We have developed standardized methods to assess a patient's ability to perform his or her treatment program. The assessment includes judgments about the patient's cognition (knowledge of key concept) and psychomotor skill (performance) with exercises and ADLs. The exercises and ADLs are those often prescribed for people with low back pain (LBP). The operational definitions and procedures proposed, however, could be applied to treatments prescribed for people with any type of neuromusculoskeletal condition. We report here the intra- and inter-rater reliability of PTs to assess independence in performance of a set of exercises and ADLs using simulated case scenarios. We hypothesize that, with training, therapists can make reliable judgments as operationally defined.

METHODS

Development process: operational definitions and procedures

The performance assessment was developed and standardized by the senior author (LVD) in collaboration with GWH and JAE. All contributors had experience treating patients with musculoskeletal pain

problems (median time 17 years, range 5.5–21 years). For this study, we chose to assess activities commonly used in the treatment of LBP (8, 9). Activity in this context refers to the therapeutic exercise or ADL being assessed. A list of the activities and key concepts that were assessed for reliability are provided in Table I. Operational definitions for activities and responses were established and the procedures for testing the 2 factors proposed to contribute to independent performance (knowledge and performance) were developed for each activity.

The first step in the development process was to decide on the key concept for each of the possible activities that could be included in a patient's treatment program. For example, the key concept to be learned for the ADL of getting in and out of bed was to avoid twisting or bending in the low back region. The second step was to decide on, and define, the possible responses for knowledge of the key concept and performance of the exercise or ADL. There were two possible responses for knowledge of the key concept; independent or dependent. A patient was independent in his or her knowledge if he or she was able to verbalize the key concept for the activity without verbal cues from the therapist. The patient was given one chance to verbalize the key concept. A person was dependent if he or she required verbal cues or demonstration of the key concept. The possible responses for performance included: (*i*) independent; (*ii*) required verbal cues; or (*iii*) required verbal cues and physical assistance. The operational definition for each response is provided in Appendix I. Table II lists the possible combinations of decisions for judgments of knowledge and performance made by the PT during the assessment.

The third step was to standardize the procedures and decision-making for assessment. To assess the patient's independence the therapist systematically reviews each activity prescribed (exercise or ADL). The PT proceeds through a series of steps to make the judgment about the patient's ability level. First, the patient is asked to perform an activity. If the patient is able to perform all aspects of the activity without verbal cues or physical assistance from the PT, the patient is judged to be independent in both knowledge and performance. The assumption of independence in knowledge is based on the proposal that

Table I. *Activities included to test rater reliability*

Exercises	Key concept†
Push up in sitting	Unweight back
Flatten low back against the wall in standing	Relax back to wall
Return from forward bending	Don't arch back; Move in hips
Hip flexor stretch in hook lying	Keep low back flat
Hip lateral and medial rotation in prone	Don't let pelvis move
Flattening lower back in sitting in a chair	Flatten back; Contract abdominals; Relax legs
Small squat in standing	Contract abdominals; Flatten back
Rock back in quadruped	Contract abdominals; Push with hands
Abdominal exercise in hook lying	Keep low back flat
Knee flexion in prone	Don't let pelvis tilt into support surface
Standing: Relax back against wall	Relax back; Don't actively push back to wall
Assume the quadruped position	Relax back down toward support surface
Hip abduction and lateral rotation in hook lying	Don't let pelvis move; Don't rotate pelvis
Hip lateral rotation in side lying	Don't let pelvis move; Don't hike pelvis
Shoulder flexion in quadruped	Don't let trunk move; Don't rotate trunk
Single leg standing in front of a table	Keep pelvis level
Hip abduction and adduction in side lying	Don't let pelvis move; Don't hike pelvis
Activities of Daily Living	Key concept
Rolling in bed	Move as unit
Assume proper sleeping position	Don't lie rotated, shifted or side bent in trunk
Assume proper sitting position	Don't sit on edge of chair; Feet must be supported
Sit to stand	Bend in hips; Don't arch back
Supine to sit	Move trunk as a unit; Don't side bend or rotate in trunk
Standing	Contract abdominals often; Feet apart; Don't stand on one leg
Stair climbing	Contract abdominals; Use handrail for support
Lifting	Squat; Lift with legs not back

*Activities refer to the therapeutic exercises or activities of daily living being assessed.
†Key concept refers to the primary goal underlying the exercise or activity of daily living and is considered important for the patient to understand in order to perform the activity.

Table II. *All possible combinations of decisions for judgments of knowledge of key concept and performance during the assessment*

Possible combinations of decisions*	Knowledge of key concept		Performance of exercise or activity of daily living			
	Independent	Dependent	Independent	Verbal cues	Verbal cues with physical assist	Unable to perform
1	×		×			
2	×			×		
3	×				×	
4	×					×
5		×		×		
6		×			×	
7		×				×

*The combination of dependent in knowledge and independent in performance has been ruled out *a priori*, based on the assumption that the patient must have knowledge of the key concept in order to perform the activity independently.

the patient must have knowledge of the key concept in order to perform the activity independently. We chose not to ask the patient to verbalize the key concept in this situation because we had observed clinically that the testing could become very repetitive. Repeated requests to verbalize the key concept would be likely to aggravate the patient. Because we made this assumption with testing, no patient would be judged as dependent in knowledge and independent in performance.

If the patient's performance is not independent the PT then asks the patient to verbalize the key concept of interest. Correct verbalization of the key concept results in a rating of independent in knowledge. If the patient cannot verbalize the key concept, the patient's knowledge is rated as dependent and the key concept is reviewed. The patient's performance is then reassessed. Since the patient has been given verbal cues related to the key concept the decision becomes whether or not the patient requires physical assistance to perform the activity. The patient is given 2 attempts to perform the activity with verbal cues. If the patient is able to perform the activity correctly, his or her rating for performance is at the verbal cue level. If the activity is not correctly performed, the PT provides physical assistance and the patient's rating of performance is at the verbal cues with physical assistance level. Appendix II is an example of the form used by the PTs to document assessment findings.

Intra- and inter-rater reliability

The procedures described were developed to assess a patient's independence in his or her treatment program during participation in a randomized clinical trial (RCT) examining outcomes of 2 conservative treatments for people with LBP. The PTs who participated in the current study were those providing treatment in the RCT. This study was approved by the Washington University Human Research Protection Office.

Examiners and training

Two PTs with experience in clinical care of people with musculoskeletal pain conditions participated in the study. One PT had 5.5 years of experience and the second PT had 21 years of experience. Training involved self-study and practical experience. The PT first studied a manual that was developed by the senior author (LVD). The manual included operational definitions for possible responses and standardized procedures for assessment and decision-making. The senior author was available for questions during the study period. A training session was provided to each PT by the senior author. The 2-h session included discussion and hands-on practice reviewing and practicing assessment of different cases. The cases were descriptions of patients who varied in their levels of knowledge and performance across a variety of exercises and ADLs. During training sessions, each therapist practiced making judgments of knowledge and performance and documented his or her judgments on a standardized assessment form. Discussion of the judgments with the senior author occurred immediately following each practice case.

Testing procedures

To assess intra- and inter-rater reliability, each PT participated in a set of standardized patient scenarios. The PTs were examined separately on 2 different occasions, with a 2-week interval between test sessions. A test session included 26 different standardized patient scenarios role-played by the senior author. The PT made judgments about knowledge and performance during each patient scenario (Appendix II). Each therapist's judgments were recorded without discussion with the examiner or the other therapist at the time of testing or during the interval between tests.

Statistical analysis

All data were analyzed using SPSS 15.0 for Windows and a custom software program written in Visual Basic (Microsoft, Inc.). Percentage of agreement (%), kappa (k) and weighted kappa (k_w) tests were used to analyze the data to examine the reliability of the therapists to make the assessments. The kappa and weighted kappa statistics are used to index therapist agreement when corrected for agreement expected by chance (10, 11). The weighted kappa is applied to ordinal data and takes into account partial agreement. The weights assigned to the 3 levels of agreement for performance assessments were as follows: (*i*) maximum agreement = 1.0; (*ii*) partial agreement = 0.50; (*iii*) maximum disagreement = 0.0.

RESULTS

The percentage agreement and kappa values to index intra-rater reliability were as follows: PT 1: % = 95 and k = 0.90 (95% confidence interval (CI) 0.70–1.00) for knowledge; % = 95 and k_w = 0.82 (95% CI 0.60–1.00) for performance and PT 2: % = 85 and k = 0.68 (95% CI 0.38–0.97) for knowledge; % = 94 and k_w = 0.80 (95% CI 0.53–1.00) for performance. The percent agreement and kappa values to index inter-rater reliability were % = 81 and k = 0.74 (95% CI 50–1.00) for knowledge, and % = 91 and k_w = 0.72 (95% CI 0.47–0.97) for performance.

DISCUSSION

In order to assess whether treatment is effective in improving a patient's outcomes, the patient must be adherent to the treatment prescribed. To be adherent to the prescribed treatment, the patient must be able to perform the treatment independently. We have described standardized methods to assess components that are important for independence; a patient's knowledge of the key concepts underlying treatment and the physical ability to perform his or her treatment. We have also demonstrated that PTs can make judgments of the patient's knowledge and performance reliably. Using the benchmarks proposed by Landis & Koch (12), trained PTs demonstrated substantial to excellent intra-
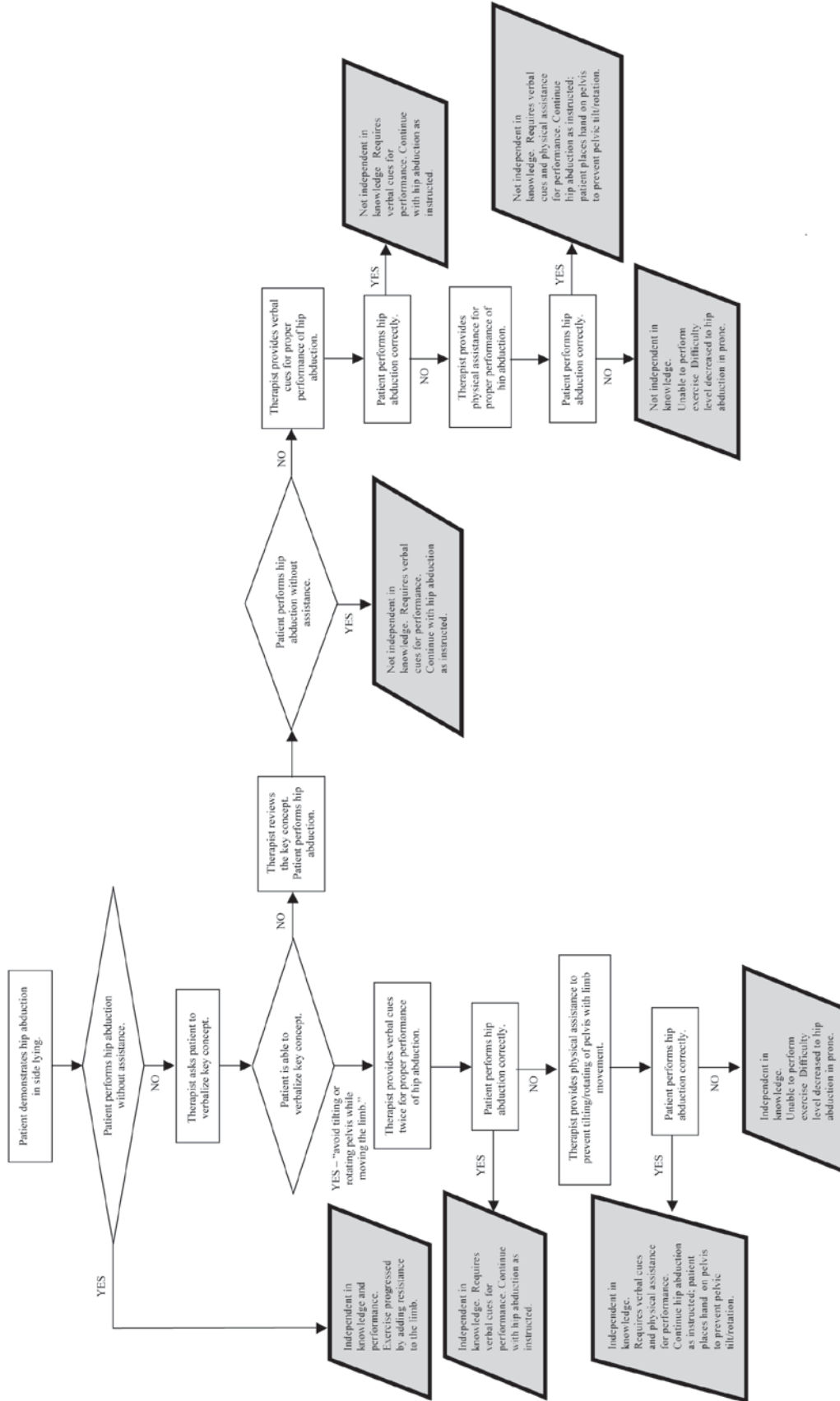
Fig. 1. Examples of the decisions and actions of the physical therapist based on the different responses of the patients.

rater reliability and substantial inter-rater reliability in assessing independence in a treatment program during standardized patient scenarios. We believe our proposed methods could be useful in the clinical and research settings.

In the clinical setting, our standardized methods can be used to determine if the patient is independent in each aspect of his or her treatment program. If a patient is not independent, the PT can use the information from the assessment to identify deficits that may result in suboptimal presentation. Specific strategies to address the identified deficits can then be used to facilitate patient independence. Fig. 1 provides an example of the decisions and actions a PT might make based on different responses demonstrated when a patient is asked to perform a prescribed strengthening exercise.

In addition to providing methods to assess independence in a prescribed exercise, our methods provide standardized procedures to assess patient performance of ADLs. Performance of ADLs is commonly assessed in patients with neuromuscular conditions using standardized instruments, such as the Functional Independence Measure (13), the Barthel Index (14) and the Modified Rankin Scale (15). We are unaware, however, of any formal assessment measures to assess ADL performance in patients with musculoskeletal pain conditions in the outpatient orthopedic setting.

It is possible that the PTs' performance assessing the activities (exercises and ADLs) included in the reliability study is not generalizable to therapist performance assessing other activities. There are 3 primary reasons we believe that the PTs' performance is likely to be generalizable. First, we tested a range of exercises and ADL items that are commonly prescribed to patients with LBP (8, 9). We included exercises that focused on: (*i*) pain relief; (*ii*) strengthening of trunk muscles; and (*iii*) trunk control. The ADL items included activities as simple as bed mobility to more difficult activities such as lifting. Secondly, the standardized patient scenarios included examples of patients who displayed a variety of levels of cognition (key concepts) and psychomotor behavior (physical performance). Finally, the therapists currently applying the measures when treating patients in our RCT have reported no difficulty making judgments of any of the exercises or ADLs prescribed.

The proposed methods for assessment of independence could be useful in future clinical treatment trials. Researchers can use the described methods to collect information about a patient's independence in his or her treatment program in conjunction with the more common methods of measuring patient adherence. We believe our methods provide a systematic assessment that will provide additional information about the patient's ability to adhere to the prescribed treatment. This additional information may provide insight into possible barriers to patient adherence and to outcomes of treatment.

The methods we have proposed are practical for the clinical and research setting. We are currently performing a RCT to compare 2 conservative treatment programs for people with chronic LBP. Thus far, the proposed methods have been applied by 4 different PTs in the treatment of 90 patients. The PTs have reported that the system does not result in additional time in treatment. They report that the system has been very

useful in formally assessing a patient's abilities and determining the specific factors preventing the patient from attaining independence. A retest of the PTs' ability to assess independence after using the assessment for one year was acceptable (unpublished data).

One factor that may have contributed positively to the rater reliability is the PTs' memory or carry-over from testing session 1 to testing session 2. To test reliability, the same standardized patient scenarios (SPSs) were used in the first and second testing sessions. It is possible that the PTs remembered the SPSs and their decisions from the first testing session. Steps were taken during the study, however, to reduce the likelihood of memory or carry-over effects. We implemented 2 strategies recommended by Sim & Wright (16). The first strategy was to present a large number of SPSs in random order. Specifically, 26 independent scenarios were used and the examiner varied the order of the SPSs from one testing session to the next. The second strategy recommended by Sim & Wright (16) was to provide a 2-week interval between the first and second testing sessions. In addition, there was no discussion of the results of the first session before the second session. Finally, at the end of the second testing session, the examiner asked each PT if he or she remembered any of the SPSs or their responses from the first testing session. Each PT responded that he or she was unable to recall his or her responses to individual SPSs. As in any study of rater reliability using a test-retest design we cannot guarantee that memory did not play a role in the therapists' reliability values. We found the SPS approach to be useful, however, because of the control of behavior variability that could be introduced with the use of actual patients with a test-retest design.

One potential limitation to our study is the use of SPSs instead of actual patients to assess rater reliability. We chose to use SPSs for 2 reasons. First, SPSs allow the examiner to provide a variety of clinical presentations that can be used across multiple testing sessions. Second, patient variability is easily controlled for using SPSs by demonstrating the same performance in each test session. To adequately test rater reliability, the patient's performance must remain stable across the testing sessions. Patient performance, however, may vary from one testing session to another due to a number of factors. In particular, a patient's performance may change due to his or her previous experience. For example, once the patient is instructed to perform an activity correctly during the first testing session, he or she may demonstrate improved performance during the second testing session. The improvement would result in different performances being assessed during the 2 testing sessions.

We believe the choice to use SPSs was appropriate for initial investigation of our standardized methods. Using SPSs is a practical and feasible method to assess rater performance that has been previously used to assess both medical student performance (17–19) and physician clinical practice (20–22). We recognize, however, that a study to assess therapists using the described system while treating actual patients would be an important addition to assessment of rater reliability.

In conclusion, using standardized patient scenarios, trained PTs displayed substantial to excellent intra-rater reliability and

substantial inter-rater reliability in assessing independence in a treatment program. Individualized treatment may be more efficient and effective if PT can make reliable judgments about the patient's knowledge of key concepts related to the treatment and performance of the treatment.

ACKNOWLEDGMENTS

REFERENCES

1. Jette AM, Smith K, Haley SM, Davis KD. Physical therapy episodes of care for patients with low back pain. Phys Ther 1994; 74: 101–110.
2. Mailloux J, Finno M, Rainville J. Long-term exercise adherence in the elderly with chronic low back pain. Am J Phys Med Rehabil 2006; 85: 120–126.
3. van Gool CH, Penninx BWJH, Kempen GIJM, Rejeski WJ, Miller GD, van Eijk JTM, et al. Effects of exercise adherence on physical function among overweight older adults with knee osteoarthritis. Arthritis Care Res 2005; 53: 24–32.
4. Lyngcoln A, Taylor N, Pizzari T, Baskus K. The relationship between adherence to hand therapy and short-term outcome after distal radius fracture. J Hand Ther 2005; 18: 2–8.
5. Kolt GS, McEvoy JF. Adherence to rehabilitation in patients with low back pain. Man Ther 2003; 8: 110–116.
6. Thomas KS, Muir KR, Doherty M, Jones AC, O'Reilly SC, Bassey EJ. Home based exercise programme for knee pain and knee osteoarthritis: randomised controlled trial. BMJ 2002; 325: 752.
7. Belza B, Topolski T, Kinne S, Patrick DL, Ramsey SD. Does adherence make a difference? Results from a community-based aquatic exercise program. Nurs Res 2002; 51: 285–291.
8. Sahrmann SA. Diagnosis and treatment on movement impairment syndromes. St Louis: Mosby, Inc.; 2002.
9. Melnik MS, Saunders HD, Saunders R. Managing back pain: daily activities guide for back pain patients. Chaska: The Saunders Group, Inc; 1998.
10. Cohen J. Coefficient of agreement for nominal scales. Educ Psychol Meas 1960; 20: 37–46.
11. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968; 70: 213–220.
12. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 1977; 33: 363–374.
13. Granger CV, Cotter AC, Hamilton BB, Fiedler RC. Functional assessment scales - a study of persons after stroke. Arch Phys Med Rehabil 1993; 74: 133–138.
14. Mahoney FI, Barthel DW. Functional evaluation: The Barthel Index. MD State Med J 1965; 14: 61–65.
15. Bonita R, Beaglehole R. Recovery of motor function after stroke. Stroke 1988; 19: 1497–1500.
16. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005; 85: 257–268.
17. Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. Acad Med 1993; 68: 443–451.
18. Rethans JJ, Saebu L. Do general practitioners act consistently in real practice when they meet the same patient twice? Examination of intradoctor variation using standardised (simulated) patients. BMJ 1997; 314: 1170–1173.
19. Beullens J, Rethans JJ, Goedhuys J, Buntinx F. The use of standardized patients in research in general practice. Fam Pract 1997; 14: 58–62.
20. Luck J, Peabody JW. Using standardised patients to measure physicians' practice: validation study using audio recordings. BMJ 2002; 325: 679.
21. Glassman PA, Luck J, O'Gara EM, Peabody JW. Using standardized patients to measure quality: evidence from the literature and a prospective study. Jt Comm J Qual Improv 2000; 26: 644–653.
22. Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. J Am Med Assoc 2000; 283: 1715–1722.

APPENDIX I. Operational definitions

Knowledge of key concept (cognition)

Operational definitions for responses to assessment of knowledge:

Independent: The patient is able to verbalize the key concept for the exercise or activity of daily living (ADL) as instructed during treatment without verbal cues from the therapist.

Dependent: The patient requires some verbal assistance or demonstration of the key concept.

Performance of activity (psychomotor skill)

Operational definitions for responses to assessment of performance:

Independent: The patient is able to perform all aspects of the exercise or ADL without physical assistance or verbal cues from the therapist. All aspects of the exercise or ADL include performance without deviations that would detract from the intended effect of the exercise or ADL. In addition, the assumption is if the patient is independent in performance he or she is also independent in his or her knowledge of the key concept. The patient would be rated as independent in knowledge of the key concept.

Verbal cues

The patient requires some verbal cues from the therapist on how to perform the exercise or ADL correctly. Given verbal cues the patient is then able to then perform all aspects of the exercise or ADL, correctly without physical assistance.

Verbal cues with physical assistance

The patient requires both: (*i*) verbal cues from the therapist on how to perform the exercise or ADL; and (*ii*) physical assistance for correct performance of all aspects of the exercise or ADL.

APPENDIX II. Treatment progress table: activities of daily living and exercises.

Activity/exercise	Review key concept	Knowledge of key concept		Performance			Visit/ date initiated	Reps (if given)
		Yes	No	I	VC	VC/ PA		
Rolling in bed **(Move as unit; Avoid rotation)**								
Sleeping position **(Don't lie with trunk rotated or side bent)**								
Sitting **(Back well supported; Don't rotate, side bend, or shift trunk)**								
Walking **(Don't rotate or hike pelvis; Shorten steps)**								
Hook lying: hip flexor stretch **(Don't let pelvis rotate; Keep pelvis still)**								
Side lying: hip lateral rotation **(Don't let pelvis move; Don't hike pelvis)**								
Prone: hip lateral and medial rotation **(Don't let pelvis move)**								

Boldface = key concept for activity of daily living or for exercise.
VC: verbal cues; PA: physical assistance; I: independent; Reps: repetitions.