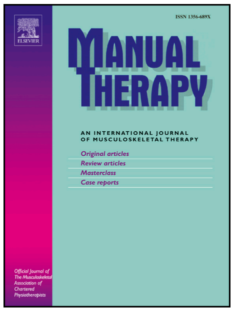

Use of Rasch analysis to investigate structural validity of a set of movement control tests for the neck

Martin Sattelmayer, Roger Hilfiker, Hannu Luomajoki, Simone Elsig

PII: S1356-689X(16)30685-3

DOI: 10.1016/j.math.2016.07.006

Reference: YMATH 1884

To appear in: *Manual Therapy*

Received Date: 19 January 2016

Revised Date: 29 June 2016

Accepted Date: 12 July 2016

Use of Rasch analysis to investigate structural validity of a set of movement control tests for the neck

Martin Sattelmayer [a], Roger Hilfiker [a§], Hannu Luomajoki [b], Simone Elsig [a]

[a] University of Applied Sciences and Arts Western Switzerland Valais (HES-SO Valais-Wallis), School of Health Sciences, Leukerbad, Switzerland

[b] Zurich University of Applied Sciences, School of Health Professions, Institute of Physiotherapy, Winterthur, Switzerland

[§]Corresponding author

Email addresses:

Martin Sattelmayer: martin.sattelmayer@gmail.com

Roger Hilfiker: roger.hilfiker@gmail.com

Hannu Luomajoki: luom@zhaw.ch

Simone Elsig: simone.elsig@gmail.com

Abstract

Background: Movement control abilities are often reduced in persons with neck pain. In physiotherapeutic practice observational tests are frequently used to assess the impaired abilities. Several tests for movement control abilities are available, but no evidence exists on how to combine and interpret them.

Objective: The aim was to investigate structural validity of a set of movement control tests with Rasch analysis.

Design: Cross-sectional study

Methods: Thirty persons with and thirty without neck pain were recruited for this study. All persons performed ten movement control tests. A partial credit model was applied to investigate item fit, ordering of the item response functions, dimensionality and hierarchy of the tests.

Results: The majority of persons with neck pain had moderate disabilities and the mean value in the Neck disability index was 10.7. Functioning of the movement control tests to measure the construct “movement control abilities” was adequate for the majority of tests. Three movement control test showed considerable misfit. Possible explanations were a reactive movement control instead of an active control and a more challenging test position. Test difficulties and person abilities could be estimated for the complete sample. The most difficult test was “sitting rocking forward” (1.13 logits) and the least difficult test was “lifting the right arm” (-1.30 logits). The highest person ability estimate was 3.61 logits indicating that movement control tests are missing to evaluate persons with moderate neck disabilities.

Conclusion: Modifying the existing set of tests is required to evaluate the complete spectrum of persons with neck pain.

Keywords:

Movement control, Neck pain, Rasch analysis

INTRODUCTION

Recently, several articles were published on the observational and non-technical assessment of the movement control capacity of persons with neck pain. Movement control capacities were defined as the ability to perform active movements while maintaining a harmonic alignment of the cervical segments with an appropriate muscle response (Elsig et al. , 2014). Movement control or proprioception is different in persons with neck pain compared to persons without neck pain (Stanton et al. , 2015).

There exist different movement control tests, either as single tests or as test batteries, for the neck: a) tests for joint position sense (e.g. relocation tests to the neutral head position (Revel et al. , 1991) or to another predefined position (Strimpakos et al. , 2006)), b) tests assessing muscle activation patterns (e.g. the craniocervical flexion test (Jull et al. , 2008) or neck extensor muscles activation (Schomacher and Falla, 2013)) c) individual observational movement control tests (Sahrmann, 2010), d) oculomotor tests (Treleaven, 2008) (e.g. the Smooth Pursuit Neck Torsion Test (Della Casa et al. , 2014)), e) tests for postural stability (e.g. the modified sensory organization test (Treleaven, 2008). Most of the studies used devices for the assessment of movement control.

Although physiotherapists often observe movement control capacities, either using non-standardized tests such as observing activities of daily living, or using standardized tests, such as tests proposed by (Sahrmann, 2010), only few studies evaluated measurement properties of observational tests. Before such tests can be recommended in practical use, they should fulfil several requirements: 1) Reliability, 2) Known group discrimination (are measures different in patients compared to healthy persons) and 3) Information should be available on how to combine different tests. There exist different possibilities how to combine individual tests: a) combination with a logistic regression formula, b) indices with statements such as "three out of ten should be positive...", c) summary scores. The summary score has several advantages, e.g. it is easy to calculate and clinicians can communicate one number. However, there are several assumptions that should be met before using a summary score.

How do observational movement control tests fulfil the requirements of reliability, known group discrimination and information on test combination? Reliability and discriminative ability for some of these was high (Elsig, 2014, Patroncini et al. , 2014, Segarra et al. , 2015). However, a summary score is frequently build by combining various observations without scientific evidence and therapists decide by pragmatic reasoning whether the observations can be combined. In item response theory, this makes only sense, when the individual tests assess the same construct (i.e. they measure a single underlying characteristic). One method to assess this is Rasch analysis (Tennant and Conaghan, 2007). Rasch analysis was introduced by George Rasch (Rasch, 1993). In contrast to classical test theory item response theory includes analysis of measurement properties on item and person level (van der Linden and Hambleton, 2013). The item level corresponds to item difficulty and the person level to person ability. In order to quantify the distance between person and item location both use the unit “logits”. One advantage of using logits is that the resulting Rasch measurement scale is interval scaled. This implies, that the distance in terms of difficulty is the same for all intervals on the scale of difficulty (i.e. one logit represents the same amount of change in difficulty irrespective of the level of difficulty). Therefore, clinicians are able to select movement control test along the measured construct according to their needs. Patients with less abilities on the construct can be measured with less difficult movement control tests and patients with more abilities can be assessed with more challenging tests. Furthermore, the administration of less tests is necessary (i.e. when the tests are administrated in ascending order with regard to their difficulty, clinicians can stop with the examination when the first test is positive, because the following tests are likely to be positive because they are more challenging).

Rasch analysis has been used in different fields, for example within the field of movement control for the development and modification of a balance test (Franchignoni et al. , 2015) or for musculoskeletal problems in a test for shoulder function (van de Water et al. , 2015). The authors reported, as benefits for clinicians, more precise measurements and a reduced amount of time needed due to a reduced number of tests without losing clinical information.

In summary, Rasch analysis may be of particular importance for clinicians because knowledge of the “difficulty” of each test allows the selection of better targeted tests. Second, with the help of person’s ability scores patients can be easily classified with regard to their abilities, which can be important for the monitoring of progression. Third, summary scores of tests can be built without the danger of mixing information of different constructs, which also implies that with the help of Rasch analysis tests can be detected which measure different abilities.

The aims of this study were to evaluate (a) the fit of the tests to the Rasch-model (i.e. to analyse if each individual test provides information about the same construct and that an overall score of the tests also refers to one single construct; (b) the relative difficulty of each test, and (c) whether the difficulty levels of the movement control tests cover the whole capacity spectrum of persons with neck pain.

METHODS

Participants

Thirty adults with neck pain (non-traumatic recurrent neck pain for more than six month) and thirty age matched persons without neck pain were recruited. Detailed information can be found in BLINDED. In short, persons with neck pain and symptoms such as feeling instability of the neck, a heavy head or aggravation of the symptoms by sustained postures were included. Excluded were patients with neurological signs, vertigo, nausea, visual disturbances, traumatic neck pain and neck surgery. The controls never felt neck or upper back pain and never had medical attention or treatment for a neck problem. Participants were recruited in a private practice setting. The study was approved by the local ethical committee and all participants signed their informed consent.

Movement control test

Ten movement control tests were used in this study. Eight of the tests showed good reliability indices (Patroncini, 2014). Two additional tests, suggested by McDonnell (Sahrmann, 2010), were included. A case control study showed good discriminative validity of three movement control tests (MC 3, MC 7 and MC 10) (BLINDED). All tests are presented in table 1.

Table 1. Included movement control tests

Test ID	Test name	Movement task
MC 1	Sitting rocking forward	Persons are instructed to bend forward with a straight upper body and back.
MC 2	Sitting rocking back	Persons are instructed to bend backward with a straight upper body and back again.
MC 3	Extension cervico-thoracic junction	Persons are instructed to make a double chin. Then, to try to look at the ceiling without losing the double chin position and without making a hollow back.
MC 4	Lifting the right arm	Persons are instructed to raise the extended arm forward and upward.
MC 5	Lifting the left arm	Persons are instructed to raise the extended arm forward and upward.
MC 6	Bilateral arm lifting	Persons are instructed to lift a weight (3kg) with straight arms to chest height (90° flexion in the shoulders) and then return with straight arms back.

MC 7	Head pro- and retraction	Persons are instructed to push the chin horizontally forward and backward.
MC 8	Supine lower neck flexion	Persons are instructed to make a double chin and then lift the entire head off the bed.
MC 9	Quadruped rocking back	Persons are instructed to make a straight back. Then to move their entire upper body backwards toward their buttocks. The whole back should remain straight during the movement.
MC 10	Quadruped cervical rotation	Persons are instructed to make a straight back. Then to turn their head and neck slowly to the right and back to the starting position (i.e. by making a rotation around an axis that runs longitudinally through their head, neck and spine). Afterwards the same movement is performed to the left side.

Overview of analyses

In order to obtain the difficulties of each test on a linear interval-level scale, we applied Rasch analyses with Winsteps, version 3. 90.0 (Linacre, 2006b). The difficulty of the tests and the abilities of the persons is reported in log of the odds, i.e. logit units. Higher logits indicate higher difficulty of the test or higher ability of the person.

Statistical model and item fit

The Partial Credit Model is a uni-dimensional latent trait model that can be used to analyse ordinal-scaled measurements (Masters, 1982, Masters and Wright, 1997). Performance of participants on each test was rated on a scale from 0 (not possible), 1 (possible after correction) and 2 (correct performance). By the means of the Rasch analysis, the ordinal rating scale is transformed to a linear interval scale with the units of logits.

Item fit

The fit of the tests to the Rasch model can be assessed with item fit statistics. These are indices, which quantify the amount to which the performance of each movement control test deviates from the expectation of the Rasch model (Bond and Fox, 2015). Linacre's guidelines were followed to investigate item fit (Linacre, 2006a). Firstly, we checked the findings for negative point-biseral correlations. Secondly, outfit mean-square statistics were investigated and lastly infit mean-scare statistics were reviewed. These goodness-of-fit statistics were evaluated according to the rule of thumb that values between 0.5 and 1.7 can be considered as reasonable fit for clinical observations (Wright et al. , 1994). When tests presented misfit on the above-mentioned statistics and misfit could not be explained by other means, the tests were removed from the pool of tests.

Metrics and ordering of the item response functions

Item response curves were plotted for all movement control tests. Metrics and ordering of the curves and corresponding item thresholds were investigated for each item response curve. In a statistical sound model the order of the item response curves should remain constant for all persons and have similar shapes (Wilson, 2004). Targeting, i.e. when the difficulty of the tests is well matched to the ability of the tested persons, was assessed by comparing test and person estimates.

Unidimensionality

Unidimensionality is the concept that only one construct of an object is assessed at a time (Bond and Fox, 2015), which is a requirement in Rasch analysis. The dimensionality was evaluated with a principal component analysis (PCA) of the standardized residuals from the Rasch analysis (Linacre, 1998). In order to diagnose a secondary dimension (i.e. reject the assumption of unidimensionality) the secondary dimension needs to have the strength of at least two tests to indicate a secondary dimension within the assessment (i.e. an Eigenvalue larger or equal than two). Furthermore, a contrast plot was searched for deviating tests and pattern to ascertain multidimensionality.

Hierarchy of the tests

Plotting of both the tests along its continuum of the difficulty and the persons along its continuum of their ability on the same axis of the logits allows the evaluation of the fit of the test difficulties to the abilities of the persons (e.g. are the tests too difficult or too easy). This allows the detection of floor- or ceiling effects.

RESULTS

Overview of the sample

The mean age of the 30 healthy persons was 37.2 years and 36.9 for the 30 persons with neck pain. In each group (healthy and persons with neck pain) were 25 women. Of the persons with neck pain twenty-four (80%) had mild disabilities as assessed with the Neck Disability Index (values below 15 points), five persons (17%) had moderate disability (values between 15 and 24) and one person (3%) had severe disability (score over 24) (Vernon, 2008). The mean value in the Neck Disability Index was 10.7 (SD: 5.12). The pain intensity at the day of the tests was 3.13 (SD: 2.01) and the pain duration: 76.7 months (SD: 78.04).

Distribution of responses

Person abilities ranged between 3.61 and -0.38 logits. A histogram of the person abilities on the movement control tests can be seen in the person item map (figure 1).

Wright map

```
MEASURE                     Person - MAP - Item
                                      |
    3             27*  9* 31  37  48  |
                                      |
                                      |
                              45  58  |
                                      |
                                      |
    2             18  26  40  46  53  |
                                      |
                                      |
                  22  49  50  52  57  |
          13* 16* 32  33  47  59  60  |
                                      |  MC_1   MC_10
    1         04* 06* 28  34  35  43  |
      12  30* 36  41* 42* 44  55  56* |  MC_2
                  01* 08* 09* 20* 23* |
                                      |  MC_7
                  03* 05* 10* 17* 38* |  MC_8
              11* 15* 19* 21  51  54* |  MC_3
    0                         24* 25* |
                                  02* |
                          07* 14* 39* |
                                      |
                                      |  MC_9
                                      |  MC_6
   -1                                 |  MC_5
                                      |
                                      |  MC_4
                                      |
                                      |
                                      |
   -2                                 |
                                      |
```

Figure 1 Wright map, *: indicates a person with neck pain. The Wright map consists of two histograms. On the left side persons are ordered with regard to their abilities (e.g. person 27 was classified as a person with excellent movement control abilities). Tests are arranged with regard to their difficulty on the right side (e.g. the movement control test MC_4 was considered as the least difficult test). All movement control tests (MC_1 to MC_10) are presented in table 1.

Item Fit

Item fit was estimated with point biserial correlations, outfit mean-square and infit mean-square statistics. All statistics are presented in Table 2.

Table 2

Overview item fit statistics; * indicates item misfit (not between 0.5 and 1.7 mean square value); # indicates tests requiring a reactive movement control

Test ID	Test name	Item difficulty estimate	Standard error	Point biserial correlation	Outfit mean-square	Infit mean square
MC 1	Sitting rocking forward #	1.13	0.18	0.55	0.95	0.83
MC 2	Sitting rocking back #	0.89	0.17	0.45	1.16	1.11
MC 3	Extension cervico-thoracic junction	0.18	0.19	0.44	0.96	1.04
MC 4	Lifting the right arm #	-1.30	0.35	0.36	0.41*	0.97
MC 5	Lifting the left arm #	-1.08	0.31	0.40	0.42*	0.95
MC 6	Bilateral arm lifting #	-0.91	0.29	0.32	0.78	0.89
MC 7	Head pro- and retraction	0.42	0.18	0.47	1.03	0.99
MC 8	Supine lower neck flexion	0.25	0.19	0.41	1.04	1.07
MC 9	Quadruped rocking back #	-0.69	0.26	0.13	1.83*	1.62
MC 10	Quadruped cervical rotation	1.10	0.18	0.62	0.93	0.98

Point-biserial correlation

All point-biserial correlations were positive, indicating that all tests worked in the intended direction. Correlations ranged between 0.13 (MC 9) and 0.62 (MC 10).

Mean square outfit and infit statistic

Three tests were identified with possible misfit on outfit mean square statistics (i.e. values under 0.5 or over 1.7). MC 9 presented potential underfit (1.83) and MC 4 (0.41) and MC 5 (0.42) indicated overfit. The remaining seven tests were appraised as presenting adequate outfit mean square fit statistics. Outfit mean square statistics are presented in Figure 2. All ten tests presented adequate fit with regard to infit mean-square fit statistics.

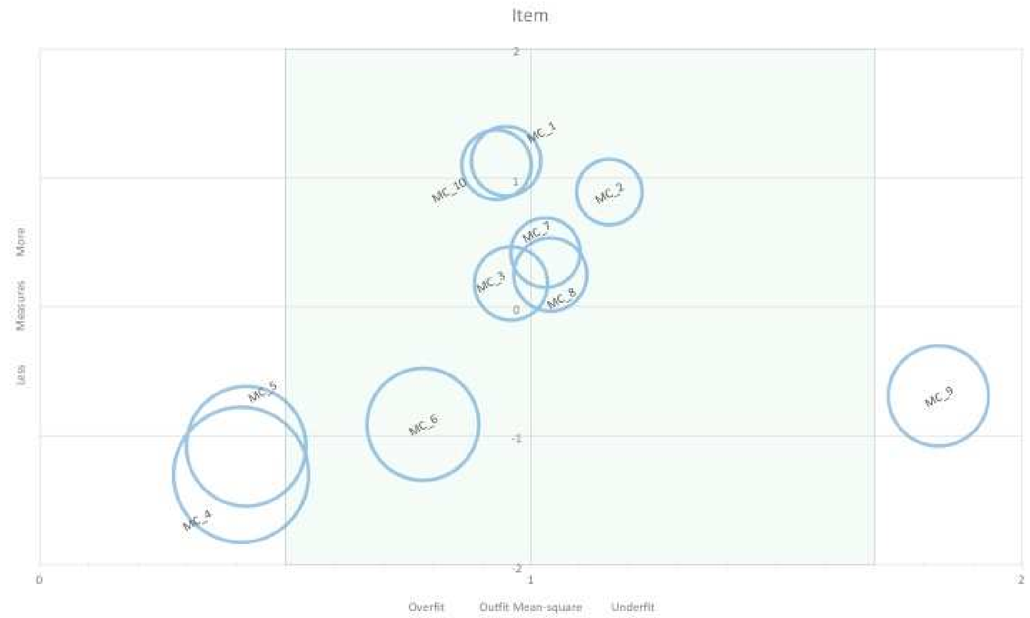

Figure 2 Bubble plot of outfit mean square statistics. Outfit mean square statistics of each movement control test are presented on the x-axis. The green area indicates adequate fit of the movement control tests to the statistical model. Tests outside this area indicate misfit. Test difficulty of the movement control tests is plotted on the y-axis. The bubble size is determined by the standard errors.

Ordering of the item thresholds

One test (MC 9) showed disordered thresholds. Within this test the difficulty estimate for a 1 was appraised as being less difficult compared to the estimate of a 0.

Unidimensionality

The PCA showed that the first residual contrast had an Eigenvalue of 2.3, this accounts for 14.9% of the unexplained variance of the first residual contrast. An Eigenvalue of 2 can be used to indicate a secondary dimension. Analysis of a contrast plot indicated two tests with a deviating pattern (MC 4 and MC 5).

Hierarchy of the tests

All person ability estimates and test difficulty estimates are illustrated on a wright map (Figure 1). Test MC 1 was analysed as having the highest test difficulty (1.13 logits) and test MC 4 as having the lowest difficulty (-1.30 logits). The person item map indicated missing tests for persons with higher ability estimates and presented evidence for a ceiling effect of the current test battery of movement control tests. The mean difficulty of the movement control tests was zero logits. Four tests were classified as being more easy and six tests as more challenging to achieve. The person item map indicated missing tests for persons with higher ability estimates (3.61 logits was the highest person ability estimate) and presented evidence for a ceiling effect of the current test battery of movement control tests. This implies that no tests were available to precisely measure persons with abilities between 1.13 and 3.61 logits.

DISCUSSION

The main finding from this investigation is that the targeting of the tests was not optimal for 14% of the patients with neck pain, i.e. there were not enough difficult tests. The tests were not able to detect movement control deficits in persons with higher abilities. In contrast, we have well targeted tests for the persons with lower abilities. However, the included sample had a moderate to low level of disability measured with the NDI. It would be possible that in a sample of persons with acute pain and high disability the targeting of the tests would be better (i.e. there would be no need for more difficult tests).

The Rasch model presented reasonable good fit to the data. However, we were not able to strongly confirm the unidimensionality of the model and some doubts remain with regard to dimensionality. The model scored slightly above the threshold of 2 Eigenvalues for unidimensionality, indicating a possible secondary dimension: Tests MC 4 and MC 5 presented a distinct pattern on a contrast plot. Both tests involve an evaluation of the neck while an arm movement is performed. Both tests were appraised as being to “simple” to challenge the equilibrium of the neck, i.e. not enough reactive moments on the trunk are created that must be countered by opposing postural moments during the movement of the arm (Huxham et al. , 2001). During MC 4, persons are asked to elevate their right arm until 90° flexion. This movement does only slightly lead to associated movements in the neck. In contrast, when the arm would be raised to 180°, there would be more associated movements in the neck. Therefore, it can be assumed that both tests would fall within the dimension “movement control tests of the neck” when they are modified accordingly.

Item fit of the movement control tests was adequate in the infit mean square statistic. Only three tests presented evidence of misfit on outfit mean-square statistics. Due to adequate infit mean-square statistics they remained in the test pool. The test MC 9 (rocking back movement in a quadruped position) was appraised as presenting underfit to the statistical model and incorrect item threshold ordering. Two issues might partly explain the deviation from the Rasch model. Firstly, this test does not involve an active movement of the neck. Instead reactive movement control capacities of the neck are evaluated. Six other movement control tests also involved reactive control and three of them presented misfit to the model. Secondly, persons are tested in a quadruped position. Therefore, a set of different movement control skills are required compared to the movement control tests 1 to 8, which are either performed in a sitting or standing position. However, MC 10 also requires a quadruped position (but involved active movement) and presented good fit to the model.

Limitations

From a clinical point of view, the included sample presents considerable heterogeneity with regard to disability and pain duration. One could argue that it would be better to analyse each subgroup separately (i.e. a specific targeting would be available for acute, subacute and chronic patient groups). This was not possible due to the low sample size. However, the current sample is representative of the persons to which the test battery can be applied to in clinical settings.

A sample size of 60 persons can be considered as a small sample for a Rasch analysis resulting in less precise estimates and less powerful fit statistics. However, the standard errors in our study were within reasonable limits for most estimates. Only standard errors of two tests (MC 4 and MC 5) were considerably larger than the remaining standard errors. Smith et al. (2008) appraised that both mean square statistics (infit and outfit) worked stable to identify misfitting tests in small samples of at least 25 persons.

Implications

The findings have three main implications for clinicians. First, estimates for the difficulties of each movement control test were established. Second, estimates for person's ability could be calculated. Therefore, clinicians are able to precisely match the administration of movement control tests to the needs of their patients. For example, if the tests are used as a screening test for impaired movement control abilities only tests with test difficulties around a pre-specified threshold value are needed in order to classify patients. If the aim is to closely evaluate a patient's progress than movement control test can be selected that precisely match the ability of the patient. Lastly, a summary score of the tests can be built without the danger of combining information of different constructs. However, because unidimensionality was not strictly confirmed in this study, two tests (MC_4 and MC_5) should not be included in the summary score.

With regard to implications for research there is a need for the design and evaluation of new more challenging movement control tests and the movement control tests MC_4 and MC_5 should be modified to better measure the underlying construct. Possibly panels with international experts can be used to respond to this challenge. Furthermore, measurement properties such as reliability indices (i.e. inter and intra-rater reliability and measurement error) and validity (e.g. known group validity) should be established for the created movement control tests. Lastly, structural validity of this new test battery (i.e. new set of tests) should be explored by the means of Rasch analysis in a considerable larger sample.

Conclusion

This investigation presented evidence of structural validity of a recently published set of movement control tests for the neck. A paucity of difficult movement control tests exits for persons with higher abilities. Increasing and modifying the existing test battery is required by further research in order to improve the use the test battery in clinical situations.

REFERENCES

Bond T, Fox CM. Applying the Rasch model: Fundamental measurement in the human sciences: Routledge; 2015.
Della Casa E, Helbling JA, Meichtry A, Luomajoki H, Kool J. Head-Eye movement control tests in patients with chronic neck pain; Inter-observer reliability and discriminative validity. BMC musculoskeletal disorders. 2014;15:16.
Elsig S, Luomajoki H, Sattelmayer M, Taeymans J, Tal-Akabi A, Hilfiker R. Sensorimotor tests, such as movement control and laterality judgment accuracy, in persons with recurrent neck pain and controls. A case-control study. Manual therapy. 2014;19:555-61.
Franchignoni F, Godi M, Guglielmetti S, Nardone A, Giordano A. Enhancing the usefulness of the Mini-BESTest for measuring dynamic balance: a Rasch validation study. European journal of physical and rehabilitation medicine. 2015;51:429-37.
Huxham FE, Goldie PA, Patla AE. Theoretical considerations in balance assessment. Aust J Physiother. 2001;47:89-100.
Jull GA, O'Leary SP, Falla DL. Clinical assessment of the deep cervical flexor muscles: the craniocervical flexion test. Journal of manipulative and physiological therapeutics. 2008;31:525-33.
Linacre JM. Detecting multidimensionality: which residual data-type works best? Journal of outcome measurement. 1998;2:266-83.
Linacre JM. Misfit diagnosis: Infit outfit mean-square standardized. Retrieved June. 2006a;1:2006.
Linacre JM. WINSTEPS Rasch measurement computer program. Chicago: Winsteps com. 2006b.
Masters GN. A Rasch model for partial credit scoring. Psychometrika. 1982;47:149-74.
Masters GN, Wright BD. The partial credit model. Handbook of modern item response theory: Springer; 1997. p. 101-21.
Patroncini M, Hannig S, Meichtry A, Luomajoki H. Reliability of movement control tests on the cervical spine. BMC musculoskeletal disorders. 2014;15:402.
Rasch G. Probabilistic models for some intelligence and attainment tests: ERIC; 1993.
Revel M, Andre-Deshays C, Minguet M. Cervicocephalic kinesthetic sensibility in patients with cervical pain. Arch Phys Med Rehabil. 1991;72:288-91.
Sahrmann S. Movement system impairment syndromes of the extremities, cervical and thoracic spines. St. Louis, Missouri: Elsevier Health Sciences; 2010.
Schomacher J, Falla D. Function and structure of the deep cervical extensor muscles in patients with neck pain. Manual therapy. 2013;18:360-6.
Segarra V, Dueñas L, Torres R, Falla D, Jull G, Lluch E. Inter-and intra-tester reliability of a battery of cervical movement control dysfunction tests. Manual therapy. 2015.
Smith AB, Rush R, Fallowfield LJ, Velikova G, Sharpe M. Rasch fit statistics and sample size considerations for polytomous data. BMC Medical Research Methodology. 2008;8:33.
Stanton TR, Leake HB, Chalmers KJ, Moseley GL. Evidence of Impaired Proprioception in Chronic, Idiopathic Neck Pain: Systematic Review and Meta-Analysis. Physical therapy. 2015.
Strimpakos N, Sakellari V, Gioftsos G, Kapreli E, Oldham J. Cervical joint position sense: an intra-and inter-examiner reliability study. Gait & posture. 2006;23:22-31.
Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Care & Research. 2007;57:1358-62.
Treleaven J. Sensorimotor disturbances in neck disorders affecting postural stability, head and eye movement control. Manual therapy. 2008;13:2-11.

van de Water AT, Davidson M, Shields N, Evans MC, Taylor NF. The Shoulder Function Index (SFInX): a clinician-observed outcome measure for people with a proximal humeral fracture. BMC musculoskeletal disorders. 2015;16:1.
van der Linden WJ, Hambleton RK. Handbook of modern item response theory: Springer Science & Business Media; 2013.
Vernon H. The psychometric properties of the Neck Disability Index. Arch Phys Med Rehabil. 2008;89:1414-5; author reply 5-6.
Wilson M. Constructing measures: An item response modeling approach: Routledge; 2004.
Wright BD, Linacre JM, Gustafson J, Martin-Lof P. Reasonable mean-square fit values. Rasch measurement transactions. 1994;8:370.