A Systematic Literature Review on the Application of Rasch Analysis in Musculoskeletal Disease — A Special Interest Group Report of OMERACT 11

Ying-Ying Leung, May-Ee Png, Philip Conaghan, and Alan Tennant

ABSTRACT. ***Objective.*** The Rasch measurement model provides robust analysis of the internal construct validity of outcome measures. We reviewed the application of Rasch analysis in musculoskeletal medicine as part of the work leading to discussion in a Special Interest Group in Rasch Analysis at Outcome Measures in Rheumatology 11.

Methods. A systematic literature review of SCOPUS and MEDLINE was performed (January 1, 1985, to February 29, 2012. Original research reports in English using "Rasch" or "Item Response Theory" in musculoskeletal diseases were assessed by 2 independent reviewers. The topics of focus and analysis methodology details were recorded.

Results. Of 212 articles reviewed, 114 were included. The number of publications rose from 1 in 1991–1992 to 23 in 2011–February 2012. Disease areas included rheumatoid arthritis (28%), osteoarthritis (16.6%), and general musculoskeletal disorders (43%). Sixty-six reports (57.9%) evaluated psychometric properties of existing scales and 35 (30.7%) involved development of new scales. Nine articles (7.9%) were on methodology illustration. Four articles were on item banking and computer adaptive testing. A majority of the articles reported fit statistics, while the basic Rasch model assumption (i.e., unidimensionality) was examined in only 57.2% of the articles. An improvement in reporting qualities with Rasch articles was noted over time. In addition, only 11.4% of the articles provided a transformation table for interval scale measurement in clinical practice.

Conclusion. The Rasch model has been increasingly used in rheumatology over the last 2 decades in a wide range of applications. The majority of the articles demonstrated reasonable quality of reporting. Improvements in quality of reporting over time were revealed. (First Release Oct 15 2013; J Rheumatol 2014;41:159–64; doi:10.3899/jrheum.130814)

Key Indexing Terms:
RASCH ANALYSIS
ITEM RESPONSE THEORY
RHEUMATOLOGY
MUSCULOSKELETAL DISORDERS

The Rasch measurement model provides a robust analysis of the internal construct validity of outcome measures[1]. Fitting data to the model, a process known as Rasch analysis, is increasingly used in health sciences and in rheumatology, with its application ranging from evaluation of the psychometric properties of existing patient-reported outcome measures, revision of observer-evaluated scales like imaging scores, and development of new instruments, to the concept of item banking and computer adaptive testing schedule (CAT)[2]. The process, which involves several components, is iterative and the recommendations for the essential components to be assessed and reported for Rasch analysis have been published[2].

Our aim was to review the application and quality of reporting of Rasch analysis in the field of rheumatology and musculoskeletal (MSK) disorders over the past 2 decades. This provides data on the acceptance of using the Rasch model in research and clinical practice, and reveals the comprehensiveness and quality of reporting of Rasch statistics over time. At the recent Outcome Measures in Rheumatology 11 (OMERACT) meeting in Pinehurst, North Carolina, USA, a Rasch Special Interest Group (SIG) was established to examine the application of the model across rheumatic diseases. This article reports the work leading to the Rasch SIG in OMERACT 11.

From the Department of Rheumatology and Immunology, Singapore General Hospital, the Duke-National University of Singapore Graduate Medical School, Singapore; Division of Musculoskeletal and Rehabilitation Medicine, University of Leeds, Leeds, UK; and the National Institute for Health Research (NIHR) Leeds Musculoskeletal Biomedical Research Unit, Leeds, UK.

Y-Y. Leung, MBChB, Department of Rheumatology and Immunology, Singapore General Hospital and Duke-National University of Singapore Graduate Medical School; M-E. Png, BSc, Department of Rheumatology and Immunology, Singapore General Hospital; P. Conaghan, MB, BS, PhD, FRACP, FRCP, Division of Musculoskeletal and Rehabilitation Medicine, University of Leeds; NIHR Leeds Musculoskeletal Biomedical Research Unit; A. Tennant, PhD, Division of Musculoskeletal and Rehabilitation Medicine, University of Leeds.

Address correspondence to Dr. Leung, Consultant,

Department of Rheumatology and Immunology, Singapore General Hospital, The Academia, level 4, 20 College Road, Singapore 169856. E-mail: katyccc@hotmail.com

MATERIALS AND METHODS

Search Strategies

A systematic review of SCOPUS and MEDLINE database from January 1, 1985, to February 29, 2012, was performed to identify English-language original research reports with keywords "Rasch" or "Item Response Theory" and "rheumatology" or "arthritis" or "back pain" or "neck pain" or "musculoskeletal." The references of all retrieved articles were also screened for potentially relevant publications.

Selection of Articles

Two reviewers (LYY and PME) independently assessed inclusion of articles, and disputes were resolved by a third reviewer (AT). We included publications that used Rasch model analysis in the evaluation of any instrument in people with MSK conditions. We excluded reports from general rehabilitation medicine and populations that included a variety of diagnoses, unless they evaluated the performance of instruments in the MSK disorders group separately or their primary aim was to compare the differential item functioning (DIF) between 2 groups, of which 1 was a defined MSK group.

Quality Assessment

Two reviewers (LYY, PME) independently scored the methodological quality of each article using a standardized checklist as described below. The checklist of quality identifiers (QI) was based on the recommendation developed to address the lack of understanding of the analytical technique of the Rasch model and its applications; and to provide guidelines on the essential points to be reported in Rasch analysis[2]. Only the presence or absence of QI was recorded in our study. Reports concerning polytomous instruments were evaluated on 10 QI, while articles on dichotomous instruments were evaluated with 9 QI. Disagreements between reviewers were identified and discussed during consensus meetings. The checklists of QI were:

1. *Stating the software for Rasch model analysis*. Most Rasch analyses were undertaken with proprietary software including WINSTEPS, RUMM, ConQuest, and others[3]. Each reports the findings in a slightly different way.
2. *Reporting the mathematical derivation of the Rasch model*. When items in a scale have 2 levels of response, the dichotomous model is chosen. When items have 3 or more levels of response, the Rasch model has at least 2 forms employing slightly different mathematics, namely the Andrich Rating Scale (AS) model or the Masters Partial Credit model[2]. The main difference is the way they handle the distance between thresholds (the probabilistic midpoint between 2 adjacent categories). The AS model expects the distance between thresholds to be equal across items. The rationale behind choice of model should be reported. In RUMM, there is a likelihood ratio statistic that helps in choosing the best polytomous model.
3. *Evaluation of the threshold ordering (for polytomous items only)*. For polytomous items, it is important to assess their category structure and ensure the responses to items are consistent with the metric estimate of the underlying construct. That is, the transition from one category to the next category should reflect and increase with the underlying latent trait being measured. When this does not happen, disordered thresholds are said to occur, and collapsing categories may be necessary. In this analysis, we assessed only whether the report assessed threshold disordering.
4. *Tests of item fit to the Rasch model*. In WINSTEPS, fit is reported as the INFIT and OUTFIT statistics; there is also a standardized fit statistics reported as ZSTD. In RUMM, chi-square statistics measure fit; they also have a residual statistic, which is the standardized sum of all difference between observed and expected values summed over all persons (very similar to the WINSTEPS OUTFIT ZSTD statistic). Although the fit statistics are not directly comparable across software packages, they take the standardized value across all persons and consequently can be similar.
5. *Tests of person fit*. A few respondents with atypical response patterns may seriously affect fit at the item level. Such response patterns would be identified by its high positive residuals. This could be related to unrecorded comorbidities, such as cognitive deficits, and removing these respondents who misfit in this way may significantly improve the internal construct validity of a scale. Thus person fit is important and some form of summary of person fit should be reported.
6. *Testing for DIF*. Previously known as item bias, DIF can affect fit to the model. This occurs when different demographic groups within the sample (e.g., younger and older, male and female) respond in a different manner to an item, despite equal levels of the underlying latent trait being measured. There may be many such different groups in each data set, but at the very least, DIF with age and sex should be assessed.
7. *Reliability*. Both WINSTEPS and RUMM have a reliability measure. In WINSTEPS, an item separation ratio is reported, and the values for group use and individual use are 1.5 and 2.5, respectively. In RUMM, an estimate of the internal consistency reliability is reported as a person separation index (PSI). Prior to version RUMM 2030, this was equivalent to Cronbach's alpha, which is expressed in the raw form; PSI is in the logit form (linear person estimate). A minimum value of 0.7 and 0.85 is required for group use and individual use, respectively. More recently, the PSI has been made sensitive to the distribution, and will diverge from alpha when the distribution of persons is skewed.
8. *Response dependency*. Response dependency is where items are linked in some way, such that the response on one item will determine the response on another. An example is where several walking items are included in the same scale. A person capable of walking 1 mile without difficulty must be able to walk 0.5 miles without difficulty. If both items (reflecting the different distances) were included in the scale, reliability would be inflated, and the measurement estimation in the Rasch model might also be affected. Dependent items can be identified through the residual correlation matrix, which should show no significant associations.
9. *Unidimensionality*. The Rasch model is a unidimensional measurement model, with the assumption that the items summed together form a unidimensional scale. There are various ways to test this. Rasch software packages usually provide a principal component analysis of the residuals. In WINSTEPS, the magnitude of the first contrast of the residual is an important indicator and generally should not be above 2. In RUMM, a series of independent t tests is conducted with the pairs of person measures fitted from 2 subsets of items identified to load positively and negatively on the first component of the residuals. Person estimates derived from the positive set of items are contrasted against those derived from the negative set. A series of individual t tests is undertaken to compare the estimates for each person. The percentage of these tests outside the range –1.96 to 1.96 should not exceed 5%. Provided the differences in estimates derived from the 2 subsets of items are normally distributed, this approach is robust enough to detect strict unidimensionality[4].
10. *Transformation table*. When an instrument fulfills the Rasch model, it can be transformed to interval scale for measurement, and a transformation table should be made available for this conversion. Equal interval scaling is important in clinical trials for accurately informing the magnitude of change and the reporting of responsiveness measures (such as minimal important differences and effect size).

RESULTS

Search Results

The search strategy resulted in 200 articles, together with 12 additional articles that were identified from the references of retrieved articles. We included a total of 114 articles in this review; the reasons for exclusion are given in Table 1. The details of included articles are summarized in Supplementary Table 1 available online at jrheum.org. Disease areas included rheumatoid arthritis (28%), osteoarthritis (OA, 16.6%), and general MSK disorders (43%). For the application of Rasch analysis, 66 (57.9%) reviewed the psychometric properties of existing instruments in ordinal

Table 1. Excluded articles and reason for exclusion (n = 98).

Reasons for Exclusion	No. of Articles (%)
Non-English	7 (7.1)
Not IRT analysis	11 (11.2)
Mixed population groups	12 (12.2)
Comments, reviews, or letters to editor	12 (12.2)
Non-musculoskeletal disorders	18 (18.4)
Not using the Rasch model	38 (38.8)
Total	98

IRT: item response theory.

scales, including cross-cultural adaptation of existing instruments. Thirty-five articles (30.7%) described the development of new instruments using the Rasch model. Among these articles, 6 described using Rasch analysis to shorten existing instruments to reduce respondent burden[5,6,7,8,9,10]. Nine articles (7.9%) aimed to teach or illustrate a methodology, usually with data originally collected for another purpose. Only a minority (4 articles) reported on the construction of an item bank or application of CAT[11,12,13,14]. The number of publications using Rasch analysis, considered in 2-year blocks, rises from 1 in 1991–1992 to 21 in 2009–2010 and 23 in the subsequent 14 months (January 1, 2011–February 29, 2012; Figure 1).

For the 5 most frequently cited articles[6,10,15,16,17], their average annual citation rates ranged from 9.78 to 18.17. The reason for high citation in 3 of these articles was the evaluation of generic instruments in the rehabilitation setting, where the instruments have wide application in a wide range of patient groups[6,10,17]. Doward, *et al* reported the development of a specific and patient-derived quality-of-life instrument for ankylosing spondylitis, the first of its kind in that condition, and thus bridged a significant gap in the literature[16]. The cross-cultural adaptation of the Dutch version of Western Ontario and McMaster Universities Osteoarthritis Index in hip OA also had a high citation rate, mainly due to the high prevalence of OA and strong Dutch centers in this area of research[15].

The Quality of Reporting

Table 2 summarizes the percentage of articles reporting our list of QI. Around 40% of the articles were using WINSTEPS or RUMM for analysis while up to 17% did not report the software they used. There was a shift from use of WINSEPS to RUMM over time. Among the instruments with polytomous item response, nearly half of the articles did not state the mathematical derivative of Rasch model chosen. Among those that reported the model, only a quarter stated the rationale behind choosing a particular model. However, the percentage reporting the Rasch model derivative and the rationale of choice has increased in recent years (Table 2).

There were 21 (18.4%) and 91 (79.8%) articles describing instruments with dichotomous and polytomous item response, and 1.8% were mixed. Among the 91 articles examining polytomous scaling and 2 articles with mixed scaling, 53.8% described threshold disordering. A higher percentage of threshold order evaluation was noted in more recent articles.

The majority of the articles reported item fit statistics while person fit statistics were reported in barely one-third. The person reliability indexes were reported in 72 articles (63.2%). Unidimensionality and item independence, which are the basic assumptions of the Rasch model, were examined in 57.2% and 42.11%, respectively. DIF or systemic bias was examined in slightly more than one-half.

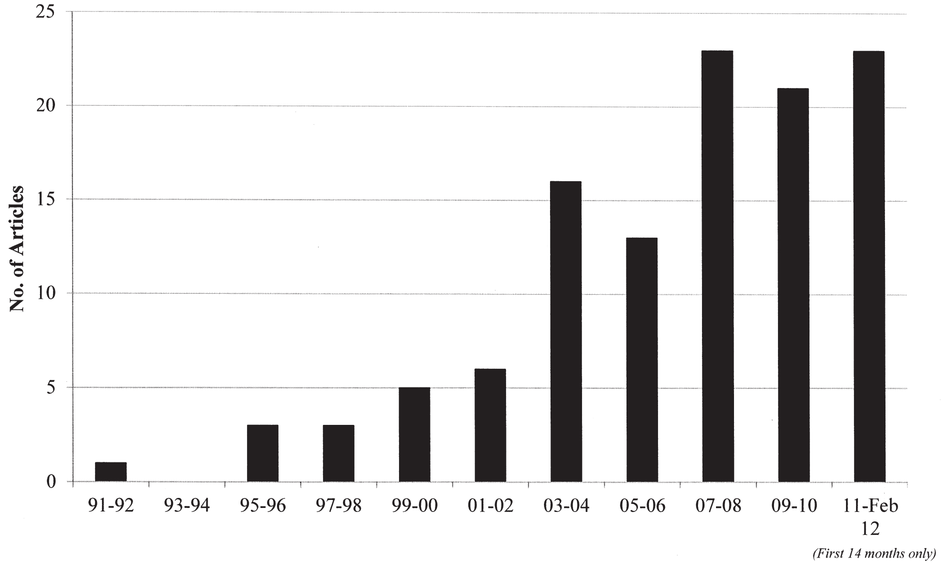

Figure 1. Number of Rasch articles, in 2-year blocks.

Table 2. Quality identifiers reported for Rasch articles (n = 114).

		No. of Articles (%)		
		All yrs (n = 114)	Before 2006 (n = 40)	From 2006 (n = 74)
1.	Software for Rasch analysis			
	Winsteps/Bigsteps	48 (42.1)	27 (67.5)	22 (29.7)
	RUMM	46 (40.4)	5 (5.0)	44 (59.5)
	ConQuest	2 (1.8)	—	2 (2.7)
	Winsteps and RUMM	1 (0.9)	—	—
	Not stated	17 (14.9)	11 (27.5)	6 (8.1)
2.	Mathematical deviation of Rasch model			
	Dichotomous item response	23	9	14
	Polytomous item response	93	32	61
	Rating Scale	14 (15.1)	2 (6.3)	12 (19.7)
	Partial Credit Model	33 (35.5)	6 (18.8)	27 (44.3)
	Both	2 (2.2)	—	2 (3.2)
	Not stated	44 (47.3)	24 (75.0)	20 (32.8)
	Rationale of choosing model stated	29 (31.2)	6 (18.8)	23 (37.7)
3.	Threshold order	50 (53.8)	8 (25.0)	42 (68.9)
	For polytomous item response (n = 93)			
4.	Item fit	111 (97.4)	39 (97.5)	72 (97.3)
5.	Person fit	44 (38.6)	7 (17.5)	37 (50.0)
6.	DIF	66 (57.9)	12 (30.0)	54 (73.0)
7.	Reliability (PIS)	72 (63.2)	14 (35.0)	56 (78.4)
8.	Response dependency	48 (42.1)	12 (30.0)	36 (48.6)
9.	Unidimensionality	65 (57.0)	9 (22.5)	56 (75.7)
10.	Transformation table	13 (11.4)	4 (10.0)	9 (12.2)

DIF: differential item functioning; PSI: person separation index.

Again, better reporting qualities were noted in more recent articles (Table 2). Overall, a higher mean number of QI were reported in more recent articles (Table 3). However, transformation tables for Rasch interval scoring were available only in 11.4%, and remained static with time.

DISCUSSION

The Rasch model has been widely used over the past 20 years to assess the quality, and help in the development, of patient-reported outcomes in health in general, and in rheumatology. Our review has shown a steady increase in use of Rasch model over time, and a concurrent steady improvement in the quality of reporting. While the most common use was either the evaluation of existing or new scales, more recent applications include the development of item banks, and the application of CAT. The majority of the articles reported item fit, but other QI were reported to a much lesser extent. Unidimensionality, which is the basic assumption of the Rasch model, was reported in only about one-half of the reports. In addition, other important QI like DIF, local independence, and reliability were reported in only about one-half of the articles. However, the field has become more sophisticated over the years, with more recent articles improving in their reporting details.

Table 3. Mean number of quality identifiers (QI) for articles in 5-year blocks.

Year of Publication	Dichotomous, n = 22 (max QI = 9)		Polytomous, n = 93 (max QI = 10)	
	No. of Articles	(Mean ± SD)	No. of Articles	(Mean ± SD)
1991–1995	0		1	(3)
1996–2000	0		10	(3.3 ± 1.3)
2001–2005	8	(3.4 ± 0.8)	21	(3.4 ± 1.9)
2006–2010	12	(4.7 ± 2.0)	40	(6.6 ± 1.9)
2011–2012 (1st 14 mo only)	2	(6.5 ± 2.1)	21	(6.9 ± 2.3)

The dissemination of the use of Rasch model has, in part, been hindered by the different Rasch software packages providing different fit statistics and different ways of assessing assumptions such as unidimensionality. This problem may increase with the advent of new packages introducing yet further fit statistics. The absence of the methodology within the mainstream statistical packages also continues to limit the uptake of the approach, including item response theory (IRT) approaches in general. This is partially reflected when the most commonly cited papers reported in our review have relatively modest average annual citation rates. However, in the context of outcomes, landmark papers such as the one introducing the Health Assessment Questionnaire have on average 70 citations a year[18], and the original paper for the Arthritis Impact Measurement Scale averages fewer than 20 citations per year[19]. Thus, given that the highest-cited Rasch paper in MSK disorders averages nearly 19 citations per year, there can be said to be progress.

Although the Rasch model has traditionally been packaged within the field of IRT, it has unique measurement properties that distinguish it from other IRT models, and lately strong arguments have been put forward that suggest the 2 approaches — Rasch Measurement and IRT — are incompatible[20]. Here, the former is seen within the framework of experimental measurement, and the latter within the framework of statistical modeling. Recently, mathematical proofs published showed that the Rasch model defines the theory of simultaneous conjoint measurement in a probabilistic framework, thus satisfying the axioms needed to produce interval scaled latent estimates[21]. Thus when a data set meets the Rasch model expectations, an interval (logit-based) estimate can be derived and this could be used whenever change scores need to be calculated from ordinal scales. Meaningful measurement is based on the arithmetical property of interval scales[22,23], and this includes the evaluation of validity and reliability; effect size, responsiveness, and minimally clinically important differences, which are the basis for "truth" and "discrimination" in the OMERACT filter, respectively[24]. This is one crucial reason for promoting the use of the Rasch model. Unfortunately, as yet, we have shown that there were only a very small number of articles that provided transformation tables to convert ordinal to interval scales for routine use in monitoring outcome and for the calculation of responsiveness and other aspects of change. The reporting of transformation tables has remained static over time, limiting the application of Rasch transformed interval measurement in clinical practice.

Looking to the future, a new direction of development for CAT can be expected in medical outcome measurement that is based on calibrated item banks (where the difficulty of items has been previously established on a single metric)[25]. It is then possible to use computer algorithms to present items to patients in such a way that their level on the construct to be measured can be determined by just a few questions[26]. This approach will greatly reduce the item load and respondent burden, with little reduction in the precision of patient ability estimates; and fulfill the "feasibility" of the OMERACT filter[24]. For example, Elhen, *et al*[11] described the application of a CAT program for assessing disability in patients with mechanical low back pain using factor analysis and Rasch analysis. The initial 108 items were identified from various instruments for disability measurement, and were calibrated onto a single matrix using data collected from 399 patients with mechanical low back pain. An exploratory factor analysis identified 2 domains, namely body function (40 items) and activity participation (54 items). The 2 domains were submitted to the Rasch model separately. The resultant item bank consisted of 33 items in body function and 49 items in activity-participation domains, which fulfilled the Rasch model of unidimensionality, item local independence, showed reliable, adequate item, and person fits, and was free of DIF with age and sex. A CAT program was developed following the logic of Thissen and Mislevy[27] and application by software, SmartCAT (v1.0). The CAT application was reevaluated in 133 patients with mechanical low back pain and found to have high correlation with the original 82-item item bank. On average only 19 items in the body function and 14 items in the activity participation were needed to estimate the precise disability levels using the CAT program.

There are several limitations in our review. First, there are many articles in the rehabilitation area that were not included in the current review. Some may have included MSK disorders, although we did include those where such conditions were clearly identified with separate results. Second, both Rasch and item response theory are not terms in the Medical Subject Headings and so we may have missed some articles if these terms do not appear in their title or abstract. Third, we assessed only the presence or absence of a limited list of QI, without examining in detail the quality of statistical analysis. Finally, we did not examine the use of Rasch transformed scales, which may be a good indicator of the application of the Rasch model.

The Rasch model has been gaining increasing acceptance in the field of rheumatology. We have shown that it has been used increasingly with improvement in reporting qualities over the past 2 decades for a wide range of applications. A few articles have demonstrated success in providing Rasch transformed scores for interval measurement, using the Rasch model in item banking and subsequent CAT application. This would be an exciting field for further development in outcome measures to improve truth, discrimination, and most of all feasibility of the OMERACT filter.

ONLINE SUPPLEMENT

Supplementary data for this article are available online at jrheum.org.

REFERENCES

1. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research, 1960, and Chicago: University of Chicago Press; 1980.
2. 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 Rheum 2007;57:1358-62.
3. Fischer GH, Molenaar IW, editors. Rasch models: foundations, recent developments, and applications. New York: Springer; 1995.
4. Smith EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002;3:205–31.
5. Katz PP, Radvanski DC, Allen D, Buyske S, Schiff S, Nadkarni A, et al. Development and validation of a short form of the valued life activities disability questionnaire for rheumatoid arthritis. Arthritis

Care Res 2011;63:1664-71.
6. Wolfe F, Michaud K, Pincus T. Development and validation of the health assessment questionnaire II: a revised version of the health assessment questionnaire. Arthritis Rheum 2004;50:3296-305.
7. Wolfe F, van der Heijde DM, Larsen A. Assessing radiographic status of rheumatoid arthritis: introduction of a short erosion scale. J Rheumatol 2000;27:2090-9.
8. Davis A, Perruccio A, Canizares M, Tennant A, Hawker G, Conaghan P, et al. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008;16:551-9.
9. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the Patient Activation Measure. Health Serv Res 2005;40:1918-30.
10. Beaton D, Wright J, Katz J. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Jt Surg Am 2005;87:1038-46.
11. Elhan AH, Oztuna D, Kutlay S, Küçükdeveci AA, Tennant A. An initial application of computerized adaptive testing (CAT) for measuring disability in patients with low back pain. BMC Musculoskelet Disord 2008;9:166.
12. Hart DL, Mioduski JE, Stratford PW. Simulated computerized adaptive tests for measuring functional status were efficient with good discriminant validity in patients with hip, knee, or foot/ankle impairments. J Clin Epidemiol 2005;58:629-38.
13. Hart DL, Cook KF, Mioduski JE, Teal CR, Crane PK. Simulated computerized adaptive test for patients with shoulder impairments was efficient and produced valid measures of function. J Clin Epidemiol 2006;59:290-8.
14. Hart DL, Mioduski JE, Werneke MW, Stratford PW. Simulated computerized adaptive test for patients with lumbar spine impairments was efficient and produced valid measures of function. J Clin Epidemiol 2006;59:947-56.
15. Roorda LD, Jones CA, Waltz M, Lankhorst GJ, Bouter LM, van der Eijken JW, et al. Satisfactory cross cultural equivalence of the Dutch WOMAC in patients with hip osteoarthritis waiting for arthroplasty. Ann Rheum Dis 2004;63:36-42.
16. Doward LC, Spoorenberg A, Cook SA, Whalley D, Helliwell PS, Kay LJ, et al. Development of the ASQoL: a quality of life instrument specific to ankylosing spondylitis. Ann Rheum Dis 2003;62:20-6.
17. Pallant JF, Tennant A. An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol 2007;46:1-8.
18. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137-45.
19. Meenan RF, Gertman PM, Mason JH. Measuring health status in arthritis. The arthritis impact measurement scales. Arthritis Rheum 1980;23:146-52.
20. Andrich D. Rating scales and Rasch measurement. Expert Rev Pharmacoeconomics Outcomes Res 2011;11:571-85.
21. Van Newby A, Conner GR, Bunderson CV. The Rasch model and additive conjoint measurement. J Appl Meas 2009;10:348-54.
22. Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. Arch Phys Med Rehabil 1989;70:857-60.
23. Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. Arch Phys Med Rehabil 1989;70:308-32.
24. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. J Rheumatol 1998;25:198-9.
25. Gershon RC. Computer adaptive testing. J Appl Meas 2005; 6:109-27.
26. Lai JS, Cella D, Chang CH, Bode RK, Heinemann AW. Item banking to improve, shorten and computerize self-reported fatigue: an illustration of steps to create a core item bank from the FACIT-Fatigue Scale. Qual Life Res 2003;12:485-501.
27. Thissen D, Mislevy RJ. Testing algorithms. In: Wainer H, Dorans N, Eignor D, Flaugher R, Green B, Mislevy R, et al, eds. Computerized adaptive testing: a primer (second edition). Hillsdale, NJ: Lawrence Erlbaum Associates; 2000:101-34.