Rheumatol Int (2003) 23: 21–26
DOI 10.1007/s00296-002-0247-2

ORIGINAL ARTICLE

Sehim Kutlay · Ayse A. Küçükdeveci
Deren Gönül · Alan Tennant

Adaptation and validation of the Turkish version of the Rheumatoid Arthritis Quality of Life Scale

Received: 25 March 2002 / Accepted: 18 August 2002 / Published online: 22 October 2002

Abstract *Objective* The aim of this study was to adapt the Rheumatoid Arthritis Quality of Life (RAQoL) questionnaire for use in Turkey and to test its reliability and validity. *Methods* The translation process included the recent guidelines for cross-cultural adaptation. Reliability of the Turkish RAQoL was assessed by internal consistency and test-retest reliability, internal construct validity by Rasch analysis, and external construct validity by associations with impairments, disability, and general health status. Cross-cultural validity was tested through analysis of differential item functioning (DIF) by comparison with data from the UK version of the RAQoL. *Results* Reliability of the adapted version was good, with high internal consistency (Cronbach's alpha 0.95 and 0.96 at times 1 and 2, respectively) and test-retest reliability (Spearman's rho 0.874). Internal construct validity was confirmed by excellent fit to the Rasch model (mean item fit 0.236, SD 1.113) and external construct validity by expected associations. The DIF for culture was found in four items. *Conclusions* Adaptation of the RAQoL for use in Turkey was successful. The instrument can be used in both national and international studies for cross-cultural comparison with the UK, as long as adjustments are made for the few items displaying DIF for culture.

S. Kutlay (✉) · A.A. Küçükdeveci · D. Gönül
Department of Physical Medicine and Rehabilitation,
Medical Faculty, Ankara University, Ankara, Turkey
E-mail: skutlay@medicine.ankara.edu.tr
Tel.: +90-312-2251898
Fax: +90-312-3094132

A. Tennant
Academic Unit of Rheumatology and Rehabilitation Medicine,
University of Leeds, Leeds, UK

S. Kutlay
Angora Evleri, Bulusmalar Caddesi, F5/25,
Beysukent, 06530 Ankara, Turkey

Keywords Adaptation · Cross-cultural · Rasch · RAQoL · Validity

Introduction

In recent years, there has been increasing interest in the assessment of outcome of chronic diseases through measurements of functional status, health status, and quality of life [1, 2, 3]. Rheumatoid arthritis (RA) is a chronic, progressive disease with high physical, social, and economic burdens and consequently has the potential for considerable effects upon quality of life [4]. The major therapeutic goals for patients with RA are to control disease activity, prevent joint deformities, preserve function, and thus maintain or improve quality of life [5].

Recently, the notion of health-related quality of life (HRQoL) has been introduced [6]. Within HRQoL, measures are generally classified into two categories: generic and condition-specific. Generic instruments are designed to capture various aspects of health status in any population, irrespective of disease or condition, thereby enabling comparison across interventions and conditions [7]. On the other hand, disease-specific quality of life instruments, as their name suggests, are targeted at a specific disease or condition and thus have the potential to be more responsive and sensitive than generic instruments [8].

Only a few generic health status measures have been adapted for use at present in Turkey [9, 10]. To date, no disease-specific measures are available with Turkish RA patients. The RA Quality of Life (RAQoL) questionnaire is a 30-item, RA-specific measure developed simultaneously in the UK and the Netherlands [11]. The U.K. English, Dutch, French-Canadian, Canadian English, Swedish, and Danish language versions have been shown to have high internal consistency, test-retest reliability, and good construct validity [11, 12, 13, 14, 15]. Consequently, the aim of the present study was to adapt the RAQoL for use in Turkey and to test the new version's reliability and validity.

Methods

The adaptation and validation of an instrument involves several stages. Initially, the *translation and adaptation process* provides an initial version of the questionnaire. An examination of *reliability* and finally *validity* usually follows. Reliability is concerned with the consistency of the instrument. Construct validity is concerned with whether the instrument measures the construct it is intended to measure. Increasingly, two forms of construct validity should be considered: *internal construct validity*, which is a more detailed examination of the structure of the scale, particularly related to unidimensionality, differential item functioning (DIF), and scaling properties [16], and *external construct validity*, which is concerned with expected associations with other key variables [17]. Finally, *cross-cultural validity* should be tested by comparison of score level attributes between the original and adapted versions [18]. The modern psychometric approach to this would be an examination of DIF by culture.

Adaptation procedure

For the translation process using the recent guidelines for cross-cultural adaptation [19], stage I involved four bilingual professionals translating the original version. Two professionals had clinical backgrounds and were thus 'informed' translators. The other two translators were an English teacher at the university and an economist educated in the UK for a master's degree, and therefore they were 'uninformed' translators. Inconsistencies in the translations were resolved (stage II) by discussions among the translators. The 'yes' and 'no' response format of the original version was retained. In order to produce meaningful and easily understandable Turkish expressions, it was necessary to change some sentence structures in the English items. Items 3 and 20 ('It's difficult to find comfortable shoes that I like' and 'My condition limits the places I can go') were translated as 'I have difficulty in finding comfortable shoes that I like' and 'Because of my condition, I am limited in the places I can go'. Similarly, items 26 and 30 ('I sleep badly at night' and 'I am limited in the clothes I can wear') were translated as 'My sleep during the night is bad' and 'I have limitations regarding the clothes I can wear'. At this stage, the accepted Turkish version of the instrument was back-translated (stage III) by a bilingual informed translator (medical doctor) who was blind to the original version. Then, all the translators and health professionals (two authors, SK and AK) involved in the adaptation process reviewed and discussed the translations and the prefinal version before field testing was considered (stage IV). Field testing for face validity (stage V) was done in a group of 25 RA patients with variable educational levels. At this stage, three items necessitated some modifications to make them more readily understandable in a Turkish setting. As the verb 'to concentrate' could not be understood by people with a low educational level, item 12 ('I find it hard to concentrate on something') was modified to 'I find it hard to keep attention on something'. Items 4 ('I avoid crowds because of my condition') and 6 ('I find it difficult to walk to the shops') were changed to 'I avoid entering crowds because of my condition'' and 'I find it difficult to walk to the shops for shopping'. This final version was then ready for further testing with respect to reliability and validity.

Patients and setting

For tests of reliability and validity, 71 patients with diagnoses of RA according to the American Rheumatism Association (ARA) criteria [20] who attend the outpatient RA clinic of a university hospital in Turkey were recruited into the study. After giving their consent to participate, all patients were assessed by the same observer (DG). The assessments included intensity of pain on a visual analogue scale (VAS), duration of morning stiffness, disease activity according to the disease activity score (DAS) [21, 22], functional disability with the Stanford Health Assessment Questionnaire (HAQ), and administration of the Turkish versions of the Nottingham Health Profile (NHP) [9] and the RAQoL. The DAS for each patient was calculated from the numbers of tender and swollen joints (both by 28-joint count), erythrocyte sedimentation rate, and patients' general health assessments by visual analogue scale. A Turkish adaptation of the HAQ had been made in Ankara University, Turkey, for a study investigating the correlation of radiological joint damage with physical disability in RA [23]. Then, validity and reliability of this Turkish version, which was also used in the present study, was documented [24]. The NHP distress index (NHPD), a 24-item single index of distress embedded within the NHP, was also calculated from the NHP [25]. The literate patients completed the questionnaires themselves, and the same observer read the items to the illiterate patients. All patients were interviewed twice, with a 2-week interval between assessments, to complete the three questionnaires (HAQ, NHP, and RAQoL).

Reliability and validity studies

Reliability

Reliability of the Turkish version of the RAQoL was tested by internal consistency and test-retest reliability. The internal consistency of an instrument is an estimate of the degree to which its constituent items are interrelated and is assessed by Cronbach's alpha coefficient [26]. Test-retest reliability is an estimate of the degree of random measurement error produced by the instrument. It is assessed by comparing instrument scores at two time points, given that there has been no change in condition between administrations [17].

Internal construct validity

The principal modern psychometric approach used to assess internal construct validity in health outcome measurement is Rasch analysis [27, 28]. The Rasch model is a unidimensional model which asserts that the easier the item, the more likely it will be passed, and the more able the person, the more likely he will pass a compared item. It assumes that the probability that a person will affirm an item or category within an item is a logistic function of the difference between the person's ability (θ) and the difficulty of the item (b) and is a function only of that difference.

$$P_i(\theta) = \frac{e^{(\theta - b_1)}}{1 + e^{(\theta - b_1)}}$$

where $P_i(\theta)$ is the probability that respondents with ability (θ) will answer item i correctly or be able to do the task specified by that item, and b is the item difficulty parameter.

From this, the expected response pattern of responses to an item set is determined, given the estimated θ and b. When the observed response pattern coincides with or does not deviate too much from the expected response pattern, then the items constitute a true Rasch scale [29]. Taken with confirmation of local independence of items, that is, no residual associations in the data after the Rasch trait has been removed, this confirms unidimensionality [30, 31].

The formulae can be expressed as a logit model:

$$\ln\left(\frac{P_{ni}}{1 - P_{ni}}\right) = \theta_n - b_i$$

In other words, the log of the odds of a yes response compared to a no response is, as stated previously, the difference between person ability and item difficulty. Thus, the Rasch analysis 'currency' is the logit (log odds unit). A logit is the distance along the line of the variable which increases the odds of observing the event by a factor of 2.718, and it is this logit scale that provides interval level measurement for data which fit the model.

The early published work on Rasch analysis in health outcome studies explored issues of unidimensionality [32], and this has remained a central theme to date [33, 34, 35]. However, Rasch analysis allows much more than an empirical test for unidimensionality. Following Lord's and Novick's work [36] and their explication of item response theory (IRT), examination of DIF became routine. The basis of the DIF approach lies in the item response function, the S-shaped trace of the proportion of individuals at the same ability level who answer a given item correctly. Under the assumption that the ability under consideration is unidimensional and that the item measures the same ability, then, except for random variations, the same curve is found, irrespective of the nature of the group for whom a function is plotted [37]. The DIF can be considered to be uniform (in which the same difference is observed across the trait) or nonuniform (in which the difference in probability between groups differs across the trait).

This analysis is thus central to issues of cross-cultural validity and, using this approach, it is possible to make a formal test of whether a scale works in the same way across cultures. Consequently, in the present study, internal construct validity was tested by fit of the data to the Rasch model and by testing DIF for age, gender, disease duration, and culture. Due to the number of repeated tests, the significance level of 0.5 was adjusted by Bonferroni correction to 0.006.

External construct validity

External construct validity is determined by testing for expected associations between the adapted instrument and other valid measures through the process of convergent construct validity [38]. In this study, associations with a range of impairments such as the duration of morning stiffness, pain intensity (VAS), and DAS were considered, as well as with functional disability as measured by the HAQ and with general health status by the NHP and NHPD.

Cross-cultural validity

Cross-cultural validity is examined by looking at the property of invariance through DIF analysis for culture. For purposes of this analysis, secondary analysis of a data set from the UK was used involving patients recruited to examine the relationship between RA and work [39]. These patients had all completed the English version of the RAQoL, and data from this version will be pooled with that of the new version to undertake the DIF analysis.

Results

Characteristics of patients

The mean age of the Turkish patients was 49.4 years (SD 14.1), and 82% were female. Forty-eight per cent were illiterate, whereas 30% had primary and 22% middle levels of education. The mean disease duration was 10.1 years (SD 7.4). The mean duration of morning stiffness was 44.3 min (SD 66.7). Mean pain intensity (VAS) was 4.98 (SD 2.95), and mean DAS was 4.29 (SD 2.27). The mean HAQ was 1.31 (SD 0.86) (time 2/retest mean 1.33, SD 0.92) and the mean RAQoL score was 17.58 (SD 9.76) (time 2/retest mean 17.39, SD 10.23).

One hundred and twelve patients were recruited in the original UK study, all of them meeting American College of Rheumatology (ACR) criteria for RA, with a mean age of 46.8 years (SD 9.99), and 75% were female. Mean HAQ was 1.16 (SD 0.74). The mean RAQoL score was 14.0 (SD 9.16).

Reliability

The internal consistency of the Turkish version of the RAQoL was excellent at both times, with Cronbach's alphas of 0.95 and 0.96 at times 1 and 2, respectively. Test-retest reliability was good, with a high correlation between the two time points (Spearman's rho 0.874).

Internal construct validity

The internal construct validity of the adapted Turkish version of the scale is confirmed by excellent fit to the Rasch model. Overall, mean item fit is 0.236 (SD 1.113) and person fit –0.072 (SD 0.797), where fit statistics are standardised to a mean of zero and standard deviation of one. No single item showed misfit to the model, with the worst-fitting item ('I try to avoid shaking hands with people') having a chi-squared value of 7.799 (df 2, $P = 0.0202$). Consequently, observed data closely follow the model expectation, and the scale constitutes a true Rasch scale. The item trait interaction chi-squared value of 85.76 (df 60, $P = 0.014$) was adequate. This shows invariance of the scale for patients at different levels of disability. Thus, the scale works in the same way across all levels of quality of life. Person separation is high at 0.942. This shows that the scale is able to discriminate at least four distinct groups of patients at different levels of quality of life.

The scale is largely free of DIF for age, gender, and disease duration. Only one item ('I find it difficult to take care of people I am close to') shows any significant difference for uniform DIF for age (Fig. 1) (Bonferroni corrected at 0.006)Fig. 1. For this item, at the same level of quality of life, younger people are more likely to say 'no', while older people are more likely to say 'yes'. Otherwise, the scale items are invariant across groups, and consequently the item response function (the S-shaped trace of expected response by level of quality of life) is identical for the different groups.

The scale has all the hallmarks of the classical ordinal scale. The thresholds are distributed in an uneven fashion across the construct, with gaps between and clusters of thresholds (Fig. 2). Patients thus lose points (i.e., improve) in a haphazard manner, depending on where they start on the scale, due to 'clustering' of the items along the metric scale. Although when data fit the Rasch model, this distortion is 'smoothed' in the middle of the scale, scores on the RAQoL from patients at the margins of the scale will be particularly affected by this distortion. Thus, data from the scale should be analysed by nonparametric statistics.

External construct validity

There were significant correlations ($P < 0.01$) between RAQoL score and duration of morning stiffness (r 0.38), VAS pain (r 0.41), and DAS (r 0.47). All these

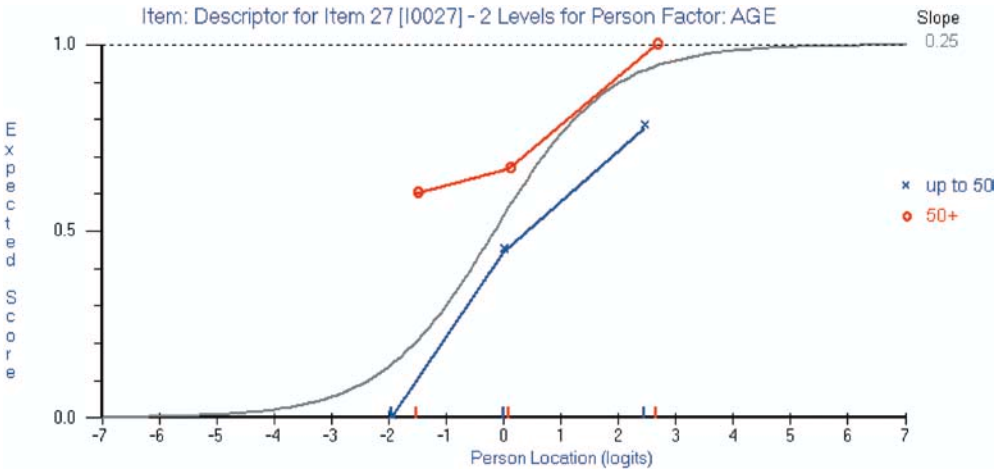

Fig. 1 Differential item functioning for the item 'I find it difficult to take care of the people I am close to' by age

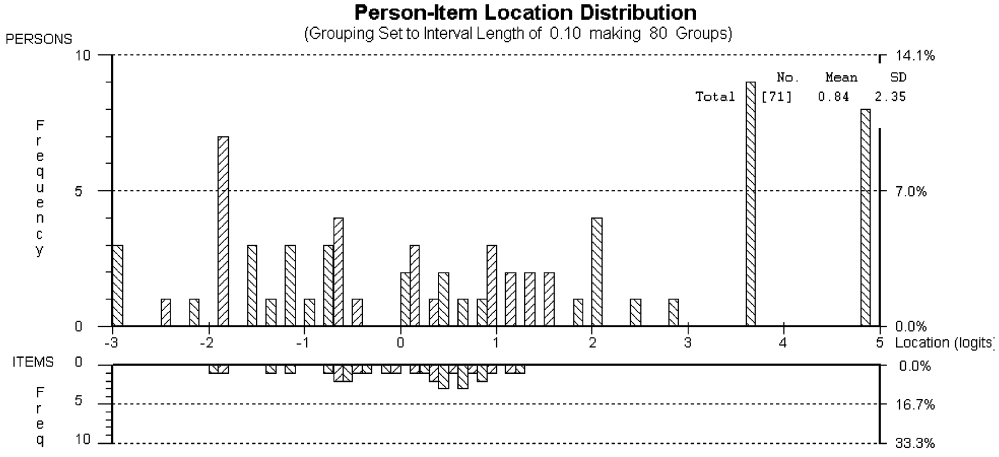

Fig. 2 Distribution of persons and items across the metric trait of quality of life

Spearman's rhos revealed moderate correlations that could be expected with impairments. The Spearman's rhos between the RAQoL and the HAQ at times 1 and 2 were 0.7 and 0.65, respectively. These higher correlations with disability compared to the impairments supported the construct validity of the Turkish version of the RAQoL, as expected. The highest correlations were found between the NHP sections of social isolation and emotional reactions and the NHPD, confirming the construct validity of the RAQoL (Table 1).

Cross-cultural validity

The cross-cultural validity of the scale is formally tested by checking the invariance of the scale across different language versions. Data from the UK data set were first fitted to the Rasch model to ensure internal construct validity. The results were very similar to the Turkish version, with good fit to the model. The mean item fit was –0.229 (SD 1.114) and person fit –0.111 (SD 0.775). Item trait interaction chi-square value is 109.3 (df 90, $P = 0.08$), showing invariance across groups of patients. Person separation was excellent at 0.941.

Invariance across countries was supported for most items in the scale, but four items showed DIF (Table 2). For these items, patients at the same level of quality of life have different probabilities of response across the two countries. For example, at the same level of quality of life, there is a significant difference in response to the item 'I find it difficult to walk to the shops'. This difference is of a uniform nature. That is, the probability of responding 'yes' to this item is different between the two countries, but the magnitude of that difference is the same across all levels of quality of life.

Also, at the same level of quality of life, those in Turkey are much more likely to respond 'yes' to the item 'I sometimes have problems using the toilet' than those in the UK (Fig. 3). Consequently, the response function

Table 1 Correlation between the RAQoL and NHP sections expressed as Spearman's rho correlation coefficient. *NHP* Nottingham Health Profile, *NHPD* NHP distress index

NHP sections	Time 1	Time 2
Energy	0.72	0.68
Pain	0.73	0.75
Emotional reactions	0.88	0.81
Sleep	0.70	0.68
Social isolation	0.87	0.83
Physical mobility	0.82	0.80
NHPD	0.92	0.83

Table 2 Items showing differential item functioning (*DIF*) by country expressed as significance level[a] for uniform (*U*) and non-uniform (*NU*) DIF

Items displaying DIF	U	NU
Finds it difficult to walk to shops	0.0006	0.0056
Sometimes has problems using the toilet	0.0000	0.0391
Often gets frustrated	0.0002	0.9812
Feels unable to control condition	0.0009	0.0500

[a] Bonferroni adjusted level of < 0.001

for those patients in Turkey, shown by the line with circles, is above the model expectation curve (that is the response function when data perfectly fit the Rasch model), while the response function for those patients from the UK, shown by the line with crosses, is below the model curve. The consistency of the difference between the two functions confirms the presence of Uniform DIF.

It is possible to adjust for the differences in response probability levels for items. Each item displaying significant DIF is allowed to be unique for each country. Thus, the 'toilet' item above becomes two items, one for each country, with responses from the other country classified as missing. This approach utilises Rasch's notion of specific objectivity, which leads to the attribute that person estimates can be made without all items completed [40]. Consequently, missing values do not obstruct making an estimate of quality of life, although the precision of that estimate will be lower. In this way, the RAQoL is expanded to 34 items, with 26 original items and the four items displaying DIF split between countries, making a further eight items, the latter displaying structural missing values for the other country responses.

Fit of the 34 items to the Rasch model within the pooled data set is good. The mean item fit was –0.275 (SD 1.397) and person fit –0.090 (SD 0.778). The item trait interaction chi-squared value is 156.1 (df 102, $P = 0.005$), showing some variance across groups of patients. However, no single item displayed misfit to the model, the worst being 'I have to go to bed earlier than I would like to' with a chi-square of 12.164 (df 3, $P = 0.0068$). Person separation was excellent at 0.944.

Discussion

Recent guidelines for cross-cultural adaptation facilitate a standardised approach to this task for all outcome measures [19]. The adaptation of the RAQoL into the Turkish language followed these guidelines. Face validity of the adapted version was confirmed at the field testing stage, which revealed that the new instrument was acceptable and relevant for Turkish RA patients. Conventional tests of reliability and validity demonstrated a successful adaptation of the RAQoL into the Turkish language similar to results for other language versions [11, 12, 13, 14, 15].

Rasch analysis showed that the adapted scale constituted a true Rasch scale and thus, fitting the data to the model provides a linear transformation of the ordinal scale. However, despite following current guidelines for cross-cultural adaptation, DIF was found to be present for four items. Given the items, this is not entirely unexpected. For example, in Turkey, squatting toilets are still common, especially in lower socioeconomic groups. Thus with a disease such as RA, it is not unreasonable to expect that, at any given level of quality of life, Turkish patients find using the toilet more difficult than their UK counterparts. The advantage of Rasch analysis is that data can be pooled and allowances made for the differences in item locations along the metric. Consequently, whereas current guidelines for adaptation may be necessary for cross-cultural adaptation, they are not a sufficient condition for cross-cultural validity, particularly when the objective is to pool data from different countries. Only the property of invariance meets this condition, and this requires that data fit the Rasch model.

In conclusion, the adaptation of the RAQoL for use in Turkey has been successful, and the instrument is

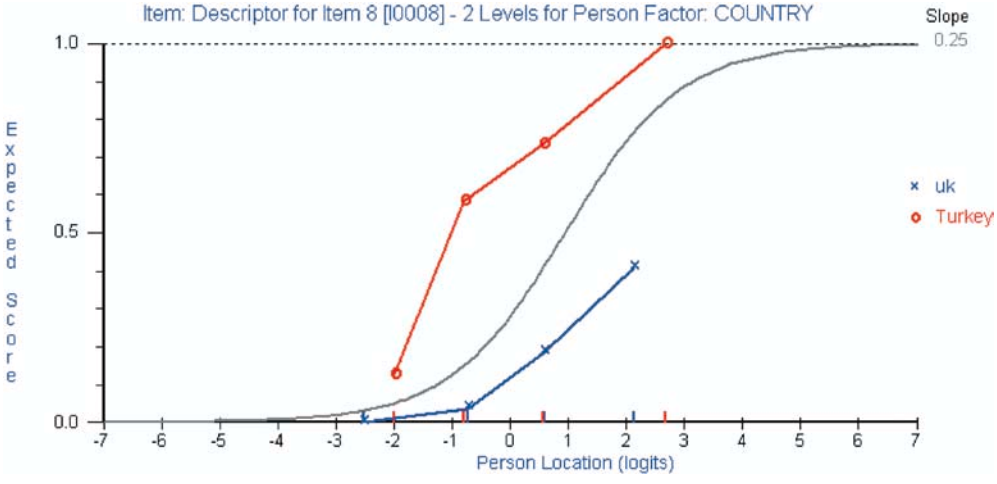

Fig. 3 Differential item functioning for the item 'I sometimes have problems using the toilet', across countries

suitable for use in clinic studies. Rasch analysis reveals that the Turkish version of the RAQoL is a robust, unidimensional ordinal measure largely free of DIF which worked well in Turkish RA patients. The instrument can also be used in international studies for cross-cultural comparison, as long as adjustments are made for the few items displaying DIF for culture.

Acknowledgement We gratefully acknowledge the support of the Scientific and Technical Research Council of Turkey.

References

1. Houssien DA, McKenna SP, Scott DL (1997) The Nottingham Health Profile as a measure of disease activity and outcome in rheumatoid arthritis. Br J Rheumatol 36:69–73
2. Guillemin F (2000) Functional disability and quality-of-life assessment in clinical practice. Rheumatol 39:17–23
3. Laupacis A, Bourne R, Rorabeck C et al (1993) The effect of elective total hip replacement on health related quality of life. J Bone Joint Surg Am 75:1619–1626
4. Whalley D, McKenna SP, De Jong Z, Van Der Heijde D (1997) Quality of life in rheumatoid arthritis. Br J Rheumatol 36:884–888
5. Lorish CD, Abraham N, Austin J, Bradley LA, Alarcon GS (1991) Disease and psychosocial factors related to physical functioning in rheumatoid arthritis. J Rheumatol 18:1150–1157
6. Shumaker S, Berzon R (1995) The International Assessment Of Health-Related Quality Of Life: theory, translation, measurement and analysis. Rapid Communications, Oxford
7. Guyatt GH, Feeny DH, Patrick DL (1993) Measuring health related quality of life. Ann Inter Med 118:622–629
8. Guyatt GH (1995) A taxonomy of health status instruments. J Rheumatol 22:1188–1190
9. Küçükdeveci A, McKenna SP, Kutlay S, Gürsel Y, Whalley D, Arasýl T (2000) The development and psychometric assessment of the Turkish version of the Nottingham Health Profile. Int J Rehabil Res 23:31–38
10. Koçyigit H, Aydemir Ö, Fişek G, Ölmez N, Memis A (1999) Kýsa form 36 (KF36)'nýn Türkçe versiyonunun güvenilirligi ve geçerliligi. Ilaç ve Tedavi Dergisi 12:102–106
11. De Jong Z, Van Der Heijde, McKenna SP, Whalley D (1997) The reliability and construct validity of the RAQoL: a rheumatoid arthritis-specific quality of life instrument. Br J Rheumatol 36:878–883
12. Neville C, Whalley D, McKenna SP, Le Comte M, Fortin PR (2001) Adaptation and validation of the Rheumatoid Arthritis Quality of Life Scale for use in Canada. J Rheumatol 28:1505–1510
13. Thorsen H, Hansen TM, McKenna SP, Sorensen SF, Whalley D (2001) Adaptation into Danish of the Stanford Health Assessment Questionnaire and the Rheumatoid Arthritis Quality of Life scale. Scand J Rheumatol 30:103–109
14. Eberhardt K, Duckberg S, Larsson BM, Johnson PM (2002) Measuring health related quality of life in patients with rheumatoid arthritis – reliability, validity, and responsiveness of a Swedish version of RAQoL. Scand J Rheumatol 31: 6–12
15. Tijhuis GJ, de Jong Z, Zwindermann AH, Zuijderduin WM, Jannsen LMA, Hazes JMW, Vlient Vlieland PM (2001) The validity of the Rheumatoid Arthritis Quality of Life questionnaire. Rheumatology 40:1112–1119
16. Tennant A, Hillman M, Fear J, Pickering A, Chamberlain MA (1996) Are we making the most of the Stanford Health Assessment Questionnaire? Br J Rheumatol 35:574–578
17. Streiner DL, Norman GR (1995) Health measurement scales. A practical guide to their development and use. Oxford University Press, Oxford
18. Ware JE Jr, Gandek B (1998) Methods for testing data quality, scaling comparisons and reliability: the IQOLA Project approach. International quality of life Assessment. J Clin Epidemiol 51:945–952
19. Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 25:3186–191
20. Arnett FC, Edworthy SM, Bloch DA et al (1988) The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 31:315–323
21. Van Riel PL (1994) Disease assessment and outcome in rheumatoid arthritis. Rheumatol Eur 23:128–129
22. Prevoo MLL, von't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LBA, van Riel PLCM (1995) Modified disease activity scores that include twenty-eight joint counts. Arthritis Rheum 38:44–48
23. Dalyan M, Küçükdeveci A, Evcik E, Ergin S (1996) Radiological joint damage in rheumatoid arthritis: relationship with certain variables. J Ankara Medical School 18:33–38
24. Küçükdeveci AA, Sahin H, Ataman S, Griffiths B, Tennant A (in press) Issues in cross-cultural validity: example from the adaptation, reliability and validity testing of a Turkish version of the Stanford Health Assessment Questionnaire (HAQ). Arthritis Care Res
25. McKenna SP, Hunt SM, Tennant A (1993) The development of a patient completed index of distress from the Nottingham Health Profile: a new measure for use in cost-utility studies. Br J Economics 6:13–24
26. Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. Psychometrika 16:297–334
27. Rasch G (1960) Probabilistic models for some intelligence and attainment tests. University of Chicago Press, Chicago
28. Andrich D (1988) Rasch models for measurement. Series: quantitative applications in the social sciences, no. 68. Sage, Newbury Park
29. van Alphen A, Halfens R, Hasman A, Imbas T (1994) Likert or Rasch? Nothing is more applicable than a good theory. J Advanced Nursing 20:196–201
30. Banerji M, Smith RM, Dedrick RF (1997) Dimensionality of an early childhood scale using Rasch analysis and confirmatory factor analysis. J Outcome Measurement 1:56–85
31. Smith RM (2000) Fit analysis in latent trait measurement models. J Appl Measurement 2:199–218
32. Silverstein B, Kilore KM, Fisher WP, Harley JP, Harvey RF (1991) Applying psychometric criteria to functional assessment in medical rehabilitation: I. Exploring unidimensionality. Arch Phys Med Rehabil 72:631–637
33. Tennant A, Geddes JLM, Chamberlain MA (1996) The Barthel Index: an ordinal score or interval level measure? Clin Rehabil 10:301–308
34. Prieto L, Alonso J, Lamarca R, Wright BD (1998) Rasch measurement for reducing the items of the Nottingham Health Profile. J Outcome Measurement 2:285–301
35. Shulman JA, Wolfe EW (2000) Development of a nutrition self-efficacy scale for prospective physicians. J Appl Measurement 1:107–130
36. Lord FM, Novick MR (1968) Statistical theories of mental test scores. Addison-Wesley, Reading
37. Angoff WH (1993) Perspectives on differential item functioning methodology. In: Holland PW, Wainer H (eds) Differential item functioning. Erlbaum, Hillsdale, pp 3–23
38. Nunnally JC (1978) Psychometric theory. McGraw-Hill, New York
39. Gillworth G, Chamberlain MA, Harvey A, Tennant A (2000) Reducing work disability in rheumatoid arthritis: development of a work instability scale. Arthritis Rheum 43 [Suppl]: S583
40. Rasch G (1977) On specific objectivity: an attempt at formalising the request for generality and validity of scientific statements. In: Glegvad M (ed) The Danish yearbook of philosophy. Munksgaard, Copenhagen, pp 58–94