Rheumatology Advance Access published April 10, 2015
RHEUMATOLOGY 53, 300
Original article doi:10.1093/rheumatology/kev008

The reliability and validity of the English version of the Evaluation of Daily Activity Questionnaire for people with rheumatoid arthritis

Alison Hammond[1], Alan Tennant[2,*], Sarah F. Tyson[3], Ulla Nordenskiöld[4], Ruth Hawkins[5] and Yeliz Prior[1]

Abstract

Objectives. The Evaluation of Daily Activity Questionnaire (EDAQ) includes 138 items in 14 domains identified as important by people with RA. The aim of this study was to test the validity and reliability of the English EDAQ.

Methods. A total of 502 participants completed two questionnaires 3 weeks apart. The first consisted of the EDAQ, HAQ, RA Quality of Life (RAQoL) and the Medical Outcomes Scale (MOS) 36-item Short-Form Health Survey (SF-36v2), and the second consisted of the EDAQ only. The 14 EDAQ domains were tested for: unidimensionality—using confirmatory factor analysis; fit, response dependency, invariance across groups (differential item functioning)—using Rasch analysis; internal consistency [Person Separation Index (PSI)]; concurrent validity—by correlations with the HAQ, SF-36v2 and RAQoL; and test–retest reliability (Spearman's correlations).

Results. Confirmatory factor analysis of the 14 EDAQ domains indicated unidimensionality, after adjustment for local dependency in each domain. All domains achieved a root mean square error of approximation <0.10 and satisfied Rasch model expectations for local dependency. DIF by age, gender and employment status was largely absent. The PSI was consistent with individual use (PSI = 0.94 for all 14 domains). For all domains, except Caring, concurrent validity was good: HAQ ($r_s = 0.72$–0.91), RAQoL ($r_s = 0.67$–0.82) and SF36v2 Physical Function scale ($r_s = -0.60$ to -0.84) and test–retest reliability was good ($r_s = 0.70$–0.89).

Conclusion. Analysis supported a 14-domain, two-component structure (Self care and Mobility) of the EDAQ, where each domain, and both components, satisfied Rasch model requirements, and have robust reliability and validity.

Key words: outcome research, rheumatoid arthritis, activities of daily living, Rasch analysis.

Rheumatology key messages

- The EDAQ evaluates the commonest everyday problems identified by people with RA.
- The EDAQ has good reliability and validity in RA and is suitable for practice and research.
- Most patients with RA considered it helpful for discussing everyday problems with an occupational therapist.

[1]Centre for Health Sciences Research, School of Health Sciences, University of Salford, *Salford,* [2]Academic Department of Rehabilitation Medicine, Section of Musculoskeletal Disease, Leeds Institute of Molecular Medicine, Faculty of Medicine and Health, The University of Leeds, Leeds, [3]Stroke & Vascular Research Centre, School of Nursing, Midwifery & Social Work, University of Manchester, Manchester, [4]Institute of Neuroscience and Physiology, Rehabilitation Medicine, Sahlgrenska Academy, Göteborg University, Sweden and [5]National Rheumatoid Arthritis Society, Derby, UK

Submitted 24 February 2014; revised version accepted 20 January 2015.

Correspondence to: Alison Hammond, Centre for Health Sciences Research (OT), University of Salford, Frederick Road, Salford M6 6PU, UK. E-mail: a.hammond@salford.ac.uk, hammond116@btinternet.com

*Present address: ICF Research Branch, Swiss Paraplegic Research, Nottwil, Switzerland

CLINICAL SCIENCE

Introduction

Although drug therapy for RA has improved, activity limitations can still occur. Commonly used patient-reported outcome measures (PROMs) of activity ability are quick and easy for patients to complete and staff to score [1–4]. However, they include few instrumental activities of daily living, insufficient detail to quantify activity limitations, have floor and ceiling effects and been devised from the perspectives of health professionals rather than of patients, which is nowadays considered poor practice [5–8]. Consequently, these PROMs are infrequently used in rehabilitation.

The Evaluation of Daily Activity Questionnaire (EDAQ) is a Swedish patient-generated PROM for RA addressing these limitations. It includes three parts: 10 numerical rating scales (NRSs) including: symptom severity, mood and life satisfaction; 11 domains assessing ability to perform daily activities, for example, eating, dressing, cooking both without (Section A) and with (Section B) ergonomic solutions (e.g. alternative methods, assistive devices); and use of assistive devices [9]. However, psychometric development is incomplete. Construct validity, floor and ceiling effects, sensitivity to change, and discriminative validity have been established with small sample sizes in Sweden [9, 10]. Test–retest reliability, concurrent validity and minimal detectable differences are unknown. Also, several scoring methods have been used [10–15], but it is unclear which is optimal.

We have developed and linguistically validated an English EDAQ to include common problematic activities of people with RA in the UK, which extended Part 2 to 138 activities in 14 domains [16, 17]. The aims of this study are to: determine the EDAQ's optimal layout and scoring method; evaluate unidimensionality of each EDAQ domain using confirmatory factor analysis (CFA); evaluate the construct validity and internal consistency of each EDAQ domain using Rasch analysis; evaluate concurrent and discriminant validity, test–retest reliability and floor and ceiling effects; evaluate sensitivity to change (using Rasch-transformed scores); and investigate acceptability, from the patients' perspective, of the EDAQ in terms of number of activities included and usefulness for discussing daily activity problems with an occupational therapist. This will provide a comprehensive, psychometrically robust standardized PROM to measure in depth the daily activity capabilities of people with RA, for both clinical and research purposes.

Methods

Participants

Ethical approval was obtained from Oldham Local Research Ethics Committee (09/H1011/25). Adults with RA attending 15 Rheumatology out-patient departments were then invited to participate by research nurses screening consecutive attendees at Rheumatology clinics (12 sites) and occupational therapists screening patients when time allowed (3 sites) according to the following criteria: a confirmed diagnosis of RA; ability to read, write and understand English; no other medical condition(s) causing difficulties performing activities of daily living; and no alteration in their disease-modifying medication regimen (which could affect test–retest reliability) in the last 3 months (and not about to). As Rasch analysis was used to assess the construct validity of Part 2, a sample size of at least 243 was recruited to give 99% confidence of the item estimate being within ±0.5 logits, irrespective of whether or not the scale was well targeted to the patients [18].

Procedures

Participants gave informed written consent, obtained according to the Declaration of Helsinki, and completed Parts 1 and 2 of the English EDAQ at home and returned it by mail. Part 3 was omitted, as the structure is not appropriate for psychometric testing. To establish optimal layout (in terms of frequency of missing data and scoring errors), two versions of Part 2 were evaluated—both the original Swedish layout and also a modified layout—with additional response options for not applicable and detailing whether assistive devices and/or help are needed. The questionnaire booklet included the evaluation of the following.

Concurrent validity

The EDAQ scores were compared with RA symptoms using the SF36v2 Bodily Pain, Vitality (fatigue) and Mental Health scales [3, 4] compared with the EDAQ Part 1 NRS. The SF36v2 was scored using Quality Metric Health Outcomes Scoring Software 4.5 [19]; activity limitations using the HAQ20 [20, 21], SF-36v2 Physical Function (PF) [3, 4] and a 0–10 hand pain NRS compared with the EDAQ Part 2 Section A scores. Quality of life was assessed using the RA Quality of Life scale (RAQoL) [22] was compared with the EDAQ Part 2 Section A scores.

Discriminant validity

EDAQ Part 2 Section A scores were compared with differing perceived health status (1 = very good through to 5 = very poor).

Acceptability of the EDAQ

Closed questions identified participants' views about whether the EDAQ would help in discussions with occupational therapists, and about the number of activities included in the EDAQ. Additionally, information to describe the recruited population was investigated, including: demographic and disease data (age, gender, marital, educational and employment status), disease duration, and RA disease-modifying medication.

Test–retest reliability

To assess this, the EDAQ was completed again 3 weeks later. Two reminders were sent at 3-week intervals if necessary.

Statistical analysis

During analysis, we determined the optimal scoring method for the EDAQ. Total scores for each domain were

created by summation of item scores within the domain, giving an indication of the person's overall ability in that domain both without (Section A) and with (Section B) ergonomic solutions. Any item not applicable in Section A was scored 0. Missing scores in Section A meant a total domain score could not be calculated. Section B scores were calculated by adding both domain item scores using ergonomic solutions and, if none were used, then that item's A score was added (as performance was unchanged). Differences between total Section A and B scores denote the impact of ergonomic solutions. (See the English EDAQ and EDAQ User Manual for scoring and examples [23, 24].)

Layout

The percentage of missing total domain scores in the original and modified EDAQ layouts were compared to identify the better format.

Unidimensionality

Data from each EDAQ domain were screened for substantive violations of unidimensionality (which may be difficult to deal with in the Rasch model framework) by CFA. In factor analysis, it is implied that observed behaviours can be described in terms of an underlying construct, and in the case of CFA, the relationship between indicators (i.e. items) and the construct is specified in advance [25]. CFA was used rather than exploratory factor analysis because the EDAQ already had a predefined domain structure. Thus, each domain was subjected to a CFA and replicated as without alternative methods, assistive devices or help (i.e. Section A), or with alternative methods or assistive devices if used (Section B). The following criteria were used for item inclusion: root mean square error of approximation (RMSEA) <0.10; Comparative Fit Index (CFI) and the Tucker Lewis Index (TLI) >0.95, and chi-square fit statistic of >0.05 (non-significant) [26]. Unidimensionality was also subsequently tested post-Rasch analysis: two estimates of the trait were derived from independent sets of items for each individual, and their difference was subjected to a *t*-test [27].

Internal consistency

Each domain was assessed using the Person Separation Index (PSI), obtained in the Rasch analysis, and Cronbach's α.

Construct validity

Construct validity was assessed using Rasch analysis [28]. The Rasch model formally tests whether data are consistent with the axioms of additive conjoint measurement from a probabilistic perspective [29]. Thus, where data are shown to fit the model expectations, an underlying quantitative structure is confirmed, and an interval scale transformation of the ordinal raw score becomes available. Data from the two layouts were analysed separately using the Partial Credit Model [30]. The process involved an iterative procedure to test stochastic ordering (fit), local independence (including response dependency and unidimensionality) and properties of invariance across groups (gender, age and employment status). Where response dependency was observed (by positive correlations of $\geqslant 0.3$ in the residuals), items were merged into testlets to absorb the dependency. Testlets were also used at the domain level to examine the possibility of higher-order constructs. Additional factors such as the validity of the structure of the polytomous items was also tested; that is, does an increasing response on an item reflect an increase in the underlying trait? Where this fails, the transition between adjacent categories (thresholds) is said to be disordered. RUMM2030 software was used for the Rasch analysis [31].

Concurrent validity

EDAQ domain scores were assessed using Spearman's correlations with measures of related constructs (as detailed above).

Discriminant validity

Discriminant validity was assessed using Kruskal–Wallis tests to evaluate differences in EDAQ scores between different perceived health status groups.

Test–retest reliability

Test–retest reliability was assessed using: weighted kappa for each item of Part 1 and each activity in Part 2 [32]; Spearman's correlations for total EDAQ domain scores; and intraclass correlation coefficients for Rasch-transformed scores [ICC (2,1)] (see below).

Sensitivity to change

Sensitivity to change was assessed by first creating Rasch-transformed scores for each EDAQ domain, conditional upon fit to the model. Acceptable levels of fit to the Rasch model are reported at the bottom of the fit table [27–29, 33]. Thus EDAQ domain raw scores were first transformed to the metric (in logits) and then linearly transformed to produce an interval-level scale of the same range for each domain. The following were then calculated: standard error of measurement (SEM); and the minimal detectable change$_{95}$ (MDC$_{95}$) score, that is, a statistical estimate of the smallest detectable change corresponding to change in ability [34, 35]. Floor and ceiling effects were considered present if $>15\%$ of participants achieved either the lowest or highest scores in any domain [36, 37].

Acceptability

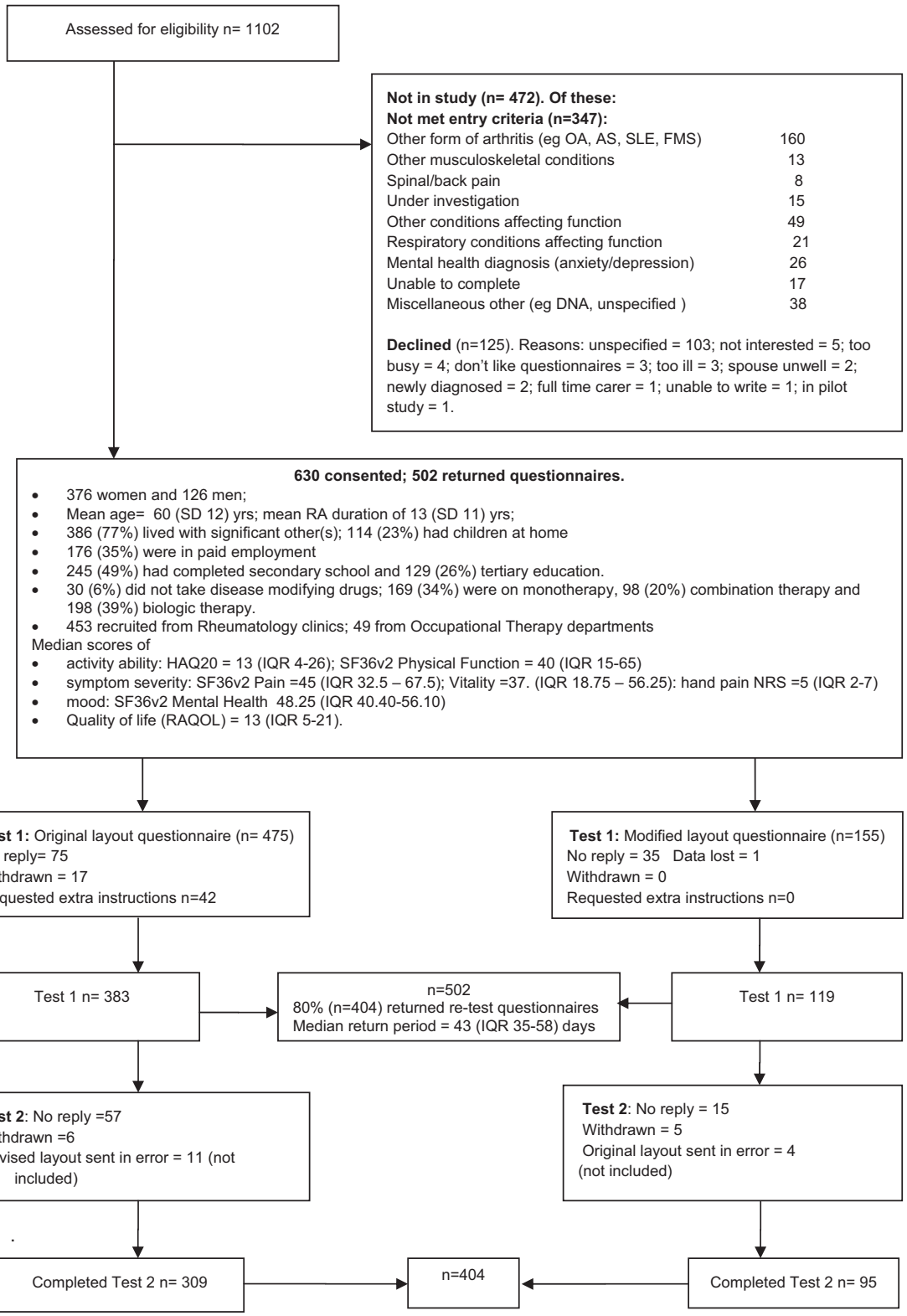
The percentages for optional responses were calculated about the EDAQ's helpfulness and number of activities included.

Results

Participants

A total of 630 participants were recruited and 404 completed all testing. Recruitment, study progress and descriptors of the participants are detailed in Fig. 1.

FIG. 1



Recruitment and study progress flow diagram

Table 1 Optimal layout and reliability testing of the EDAQ Part 2 ($n = 502$)

EDAQ domain (score range)	Original layout, n (%) missing domain scores ($n = 383$)	Revised layout, n (%) missing domain scores ($n = 119$)	Test 1 A α	Test 1 Section A Score ($n = 502$)	Test 1 Section B score ($n = 502$)	Test 2 Section A score	Test 2 Section B score	Test–retest Section A r_s	Test–retest Section B, r_s
1. Eating and drinking (0–33)	26 (6.78)	5 (4.20)	0.95	8 (4–15)	6 (2–12)	8 (4–14)	6 (2–11)	0.83*	0.83*
2. In the Bathroom and personal care (0–36)	43 (11.23)	7 (5.88)	0.93	3 (1–7)	3 (1–7)	3 (1–7)	2 (0–7)	0.86*	0.85*
3. Dressing (0–33)	31 (8.09)	5 (4.20)	0.95	6 (2–13)	6 (1–13)	5 (1–12)	5 (1–11)	0.84*	0.83[a]
4. Bathing and showering (0–33)	13 (3.39)	3 (2.52)	0.94	7 (2–15)	6 (2–14)	6 (2–13)	5 (1–12)	0.88*	0.88*
5. Cooking (0–42)	31 (8.09)	5 (4.20)	0.96	9 (3–18)	8 (2–17)	8 (2–17.5)	7 (1–16)	0.88*	0.86*
6. Moving indoors (0–36)	50 (13.05)	8 (6.72)	0.95	9 (4–15.75)	8 (3–15)	8 (3–15)	7 (2–15)	0.87*	0.86*
7. Cleaning the house (0–27)	18 (4.70)	3 (2.52)	0.94	7 (2–14)	7 (2–14)	6 (2–13)	6 (2–12)	0.85*	0.85*
8. Laundry and clothes care (0–27)	16 (4.18)	3 (2.52)	0.94	5 (1–13)	5 (1–12)	4 (1–12)	4 (1–11)	0.86*	0.85*
9. Moving and transfers (0–18)	10 (2.61)	1 (0.84)	0.90	4 (1–7)	4 (1–7)	3 (1–6)	3 (1–6)	0.83*	0.82*
10. Communication (0–18)	17 (4.44)	3 (2.52)	0.89	2 (0–5)	2 (0–4.25)	2 (0–4)	1 (0–4)	0.84*	0.81*
11. Moving outdoors and shopping (0–39)	42 (10.97)	12 (10.08)	0.94	9 (4–17)	8 (3–16)	7 (2–16)	7 (2–15)	0.89*	0.88*
12. Gardening and house maintenance (0–21)	17 (4.44)	5 (4.20)	0.94	7.5 (2–16)	7.5 (2–16)	7 (2–15)	7 (1.5–15)	0.72*	0.72*
13. Caring ($n = 275$) (0–27)	22 (5.74)	3 (2.52)	0.97	0 (0–3)	0 (0–3)	0 (0–1)	0 (0–1)	0.58*	0.58*
14. Leisure and social activities (0–27)	28 (7.31)	7 (5.88)	0.90	4 (1–9)	3 (1–9)	3 (1–8)	3 (1–8)	0.71*	0.70*

Comparison of missing domain scores for the Original ($n = 383$) and Modified ($n = 119$) EDAQ layouts; internal consistency (Cronbach's α); median (IQR) Sections A and B total domain scores and test–retest reliability (Spearman's correlations). *$P \leqslant 0.001$; α = Cronbach's alpha; r_s = Spearman's correlation coefficient.

Optimal layout

The modified layout had fewer missing total scores for most domains except Moving outdoors, and Gardening and household maintenance (Table 1). Removing items in these domains did not improve completeness of data, thus all were retained. There were no significant differences between the demographic or disease characteristics or EDAQ scores of participants who completed the two different layouts. Data for the two layouts were analysed separately for Rasch analysis and combined for classical testing ($n = 502$).

Unidimensionality, internal consistency and Rasch transformation scores

As results from the two layouts were similar, data for the original layout only were used in the CFA and Rasch analysis. CFA of the 14 domains for Part 2 Section A indicated the potential for unidimensionality after adjustment for local dependency (correlated errors) within each domain. All domains achieved RMSEA < 0.10, and CFI and TLI > 0.95 (Table 2). Nevertheless, none produced a non-significant chi-square fit, indicating some problems remained in the data. After adjustments (using testlets), all domains satisfied the Rasch model expectations for local dependency (Table 2). Virtually all items had ordered thresholds, and where this was not the case, the disordering was not statistically significant (i.e. threshold estimates had overlapping confidence intervals) (e.g. supplementary Fig. S1, available at *Rheumatology* Online). DIF by age, gender and employment status was also largely absent, with a few exceptions. For example, in Gardening and household maintenance, the item Climbing ladders showed significant DIF by age, with older people showing greater problems. All individual domains supported strict unidimensionality.

Most domains had good internal consistency, showing a PSI consistent with individual use, even after adjustment

Table 2 Rasch analysis summary for the EDAQ Section A: without alternative methods, assistive devices or help

Analysis No.	Domain	Item residual, mean (S.D.)	Person residual, mean (S.D.)	Chi-square interaction value (df)	*P*-value	PSI	Unidimensionality, % *t*-test (CI)
1	Eating	0.200 (0.538)	−0.491 (1.023)	17.41 (15)	0.295	0.89	7.1 (4.8, 9.5)
2	Personal care	−0.702 (1.805)	−0.604 (0.921)	14.56 (15)	0.483	0.77	1.6 (-0.1, 4.4)
3	Dressing	−0.097 (0.641)	−0.494 (1.090)	15.83 (28)	0.727	0.84	3.1 (0.1, 5.5)
4	Bathing	−0.338 (1.260)	−0.395 (1.026)	38.82 (25)	0.091	0.85	4.8 (1.9, 6.6)
5	Cooking	−0.007 (1.270)	−0.405 (0.934)	22.39 (20)	0.320	0.90	3.4 (0.9, 5.8)
6	Moving indoors	−0.635 (2.684)	−0.427 (1.006)	36.57 (20)	0.013	0.87	3.2 (0.8, 5.6)
7	Cleaning	−0.617 (1.252)	−0.394 (0.880)	62.14 (45)	0.046	0.88	5.9 (3.6, 8.2)
8	Laundry	0.053 (0.753)	−0.469 (1.056)	22.81 (20)	0.298	0.83	4.3 (1.6, 6.9)
9	Transfers	0.028 (1.578)	−0.385 (1.041)	31.99 (25)	0.158	0.80	2.8 (0.4, 5.2)
10	Communication	0.011 (1.205)	−0.324 (0.834)	31.18 (30)	0.407	0.75	2.7 (0.0, 5.3)
11	Moving outdoors	0.423 (1.345)	−0.402 (1.038)	36.29 (25)	0.067	0.84	4.4 (2.1, 6.7)
12	Garden and household maintenance	−0.194 (0.945)	−0.466 (0.670)	12.06 (10)	0.28	0.91	1.9 (−1.0, 4.9)
13	Caring	−0.411 (1.358)	−0.948 (2.050)	43.34 (45)	0.542	0.84	5.2 (1.2, 9.2)
14	Leisure and social activities	−0.073 (1.264)	−0.704 (1.291)	20.55 (20)	0.424	0.31	3.8 (0.4, 7.2)
15	14 domains	−0.199 (2.205)	−0.355 (1.116)	75.44 (70)	0.307	0.94	10.0 (7.8, 12.3)
16	Self-care component	−0.517 (1.961)	−0.434 (1.045)	28.72 (35)	0.764	0.91	4.5 (2.1, 6.8)
17	Mobility component-7	−0.357 (1.099)	−0.588 (1.062)	23.45 (45)	0.932	0.87	5.3 (2.8, 7.7)
18	Mobility component-5	−0.275 (1.222)	−0.500 (1.028)	16.57 (25)	0.897	0.88	4.0 (1.6, 6.5)
	Fit criteria	0.0 <1.4[a]	0.0 <1.4	> 0.05[b]		>0.85	Lower CI < 5%

[a]Where testlets are used, this may be inflated. [b]Bonferroni adjusted (average is ~0.005).

for local dependency. Where values were lower (e.g. Table 2, analysis 2), this was largely a consequence of skewed data and the presence of a substantive floor effect, which affects the PSI. This was also tested using Cronbach's α and was good to excellent at 0.89 or higher for all domains (Table 2). When each domain was treated as a testlet, and summed together to give a total score, multidimensionality was seen (Table 2, analysis 15) in two components: Self care (domains: 1, 2, 3, 5, 7, 8 and 10) and Mobility (domains: 4, 6, 9, 11, 12, 13 and 14). The items (testlets) within these components showed satisfactory fit to the Rasch model, including strict unidimensionality (Table 2, analyses 16 and 17). The CFA was repeated and again failed the chi-square test but satisfied the RMSEA < 0.10, and CFI and TLI > 0.95 criteria. Extensive modifications for local dependency were indicated within each item set.

For the Self care component, the average for persons was below the scale average (Fig. 2), indicating our sample had fewer problems with Self care activities than another sample perfectly targeted at the scale. In contrast, the Mobility component was slightly better targeted, with the mean of persons (in mobility) at −0.60 logits closer to the mean of the items (Fig. 2). The analysis was then re-run with data from Section B, which, when present (i.e. alternative methods or assistive devices were used), overrode Section A responses. The results were almost identical to those of Section A, including those of the Self care and Mobility components. A transformation table to convert ordinal EDAQ Part 2 scores to interval data is provided in the EDAQ User Manual [24].

Following classical testing, the Caring and Leisure and social activities domains were removed from the Mobility component, as there were large numbers of missing items. Fit of the remaining domains to the Rasch model within the revised Mobility component was good (Table 2, analysis 18), and the Rasch transformation table was adjusted.

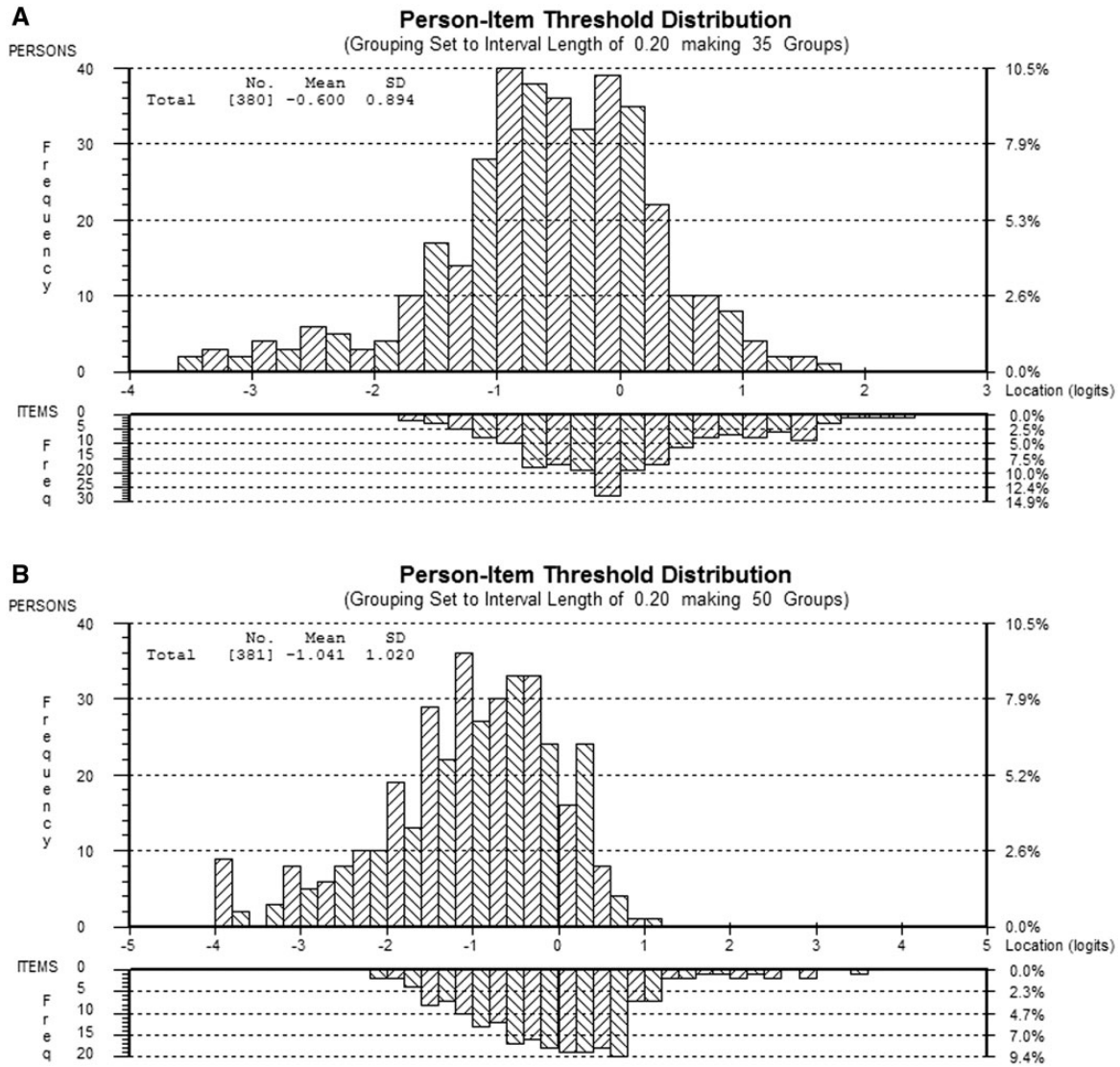
Concurrent validity

Strong correlations were found between: EDAQ Part 1 (NRS) and relevant SF36v2 scales ($r_s = -0.62$ to -0.76, Table 3); and EDAQ Part 2 and HAQ20 (disability) scores ($r_s = 0.72$–0.91), the RAQoL ($r_s = 0.67$–0.82), the SF36v2 PF scale ($r_s = -0.60$ to -0.84), hand pain NRS ($r_s = 0.55$–0.70) and SF36v2 Pain ($r_s = -0.54$ to -0.69), apart from the Caring domain, which showed weak to moderate correlations ($r_s = -0.25$ to -0.32; Table 4). The SF36v2 Vitality scale was moderately correlated with all EDAQ domains ($r_s = -0.45$ to -0.57), apart from Caring, which was weak ($r_s = -0.25$) (Table 4).

Discriminant validity

There were significant differences in all EDAQ domain scores ($P < 0.000$) between participants with very

FIG. 2



Distribution of persons and item thresholds for (**A**) the Self care and (**B**) the Mobility components

good/good ($n = 177$), fair ($n = 191$) and poor/very poor ($n = 128$) health.

Test–retest reliability

For each of the Part 1 10 NRS this was moderate (linear weighted kappa scores = 0.41–0.58) (Table 3) and good for Part 2 Sections A and B ($r_s = 0.70$–0.89), apart from domain 13 (Caring), which had moderate reliability ($r_s = 0.58$) for both sections (Table 1). ICC (2,1) for Rasch-transformed Part 2 EDAQ domains were 0.70–0.89, apart from domain 13 (Caring), which was 0.62. Linear weighted kappa scores for individual items in EDAQ Part 2 domains ranged from 0.35 to 0.75. For Section A, 2 items had fair, 58 had moderate and 78 had good agreement. For Section B, 2 had fair, 69 had moderate and 67 had good agreement. The two items with only fair agreement were: use a computer and mouse (Communication) and feed a child (Caring).

Sensitivity to change

The SEM ranged from 0.62 to 3.09 and MDC_{95} from 1.71 to 4.86 in the 11 domains of the original Swedish EDAQ, but was larger for the three new domains (Table 4).

Floor and ceiling effects

Six domains demonstrated floor effects: Bathroom and personal care ($n = 100$; 19.9%); Laundry and clothes care ($n = 82$; 16.3%); Moving and transfers ($n = 90$; 17.9%); Communication ($n = 160$; 31.9%); Caring ($n = 277$; 55.2%); and Leisure ($n = 79$; 15.7%). No ceiling effects occurred.

Table 3 Psychometric testing of EDAQ Part 1 (numeric rating scales) ($n = 502$)

Numerical rating scale (0–10)	Test 1, median (IQR)	Test 2, median (IQR)	Test–retest reliability, linear weighted kappa	Correlations with SF36v2
Disease activity	5 (3–7)	4 (3–6)	0.48	
Mood	4 (2–5)	4 (2–5)	0.41	−0.62[a,*]
Pain when resting	3 (2–6)	3 (1–5)	0.53	-0.69[b,*]
Pain when moving	5 (3–7)	5 (2–7)	0.54	-0.77[b,*]
Stiffness	5 (3–7)	5 (3–7)	0.52	
Joint movement limitations	5 (3–7)	5 (3–7)	0.57	
Fatigue	6 (4–8)	6 (3–7)	0.54	-0.70[c,*]
Worry	4 (2–6)	4 (2–6)	0.50	
Sleep	5 (2–7)	4 (2–7)	0.58	
Satisfaction with life	5 (2–7)	5 (2–6)	0.46	

[a]Short form F36v2 Mental Health scale. [b]Short form 36v2 Bodily Pain scale. [c]Short form 36-v2® Vitality scale. *$P \leqslant 0.001$.

Table 4 Psychometric testing of EDAQ Part 2 section A scores ($n = 502$)

EDAQ domain (score range)	HAQ20	SF36v2: PF	SF36v2: Vitality	SF36v2: Pain	Hand pain	RAQOL	Perceived Health	ICC (2,1)	SEM	MDC_{95}
1. Eating (0–33)	0.83*	−0.70*	−0.48*	−0.60*	0.70*	0.73*	0.62*	0.82 (0.79, 0.85)	1.70	4.70
2. In the Bathroom and personal care (0–36)	0.86*	−0.73*	−0.52*	−0.60*	0.66*	0.77*	0.63*	0.87 (0.84, 0.89)	1.26	3.48
3. Dressing (0–33)	0.86*	−0.73*	−0.52*	−0.64*	0.67*	0.74*	0.64*	0.85 (0.82, 0.88)	1.76	4.86
4. Bathing and showering (0–33)	0.88*	−0.76*	−0.53*	−0.63*	0.66*	0.78*	0.62*	0.88 (0.86, 0.91)	0.96	2.65
5. Cooking (0–42)	0.89*	−0.77*	−0.54*	−0.66*	0.73*	0.80*	0.68*	0.88 (0.85, 0.90)	1.72	4.75
6. Moving indoors (0–36)	0.90*	−0.84*	−0.55*	−0.69*	0.69*	0.82*	0.66*	0.79 (0.75, 0.83)	1.38	3.81
7. Cleaning the house (0–27)	0.86*	−0.76*	−0.53*	−0.64*	0.68*	0.76*	0.66*	0.85 (0.81, 0.87)	1.31	3.62
8. Laundry and clothes care (0–27)	0.86*	−0.74*	−0.50*	−0.64*	0.70*	0.77*	0.64*	0.86 (0.84, 0.89)	1.27	3.51
9. Moving and transfers (0–18)	0.86*	−0.75*	−0.57*	−0.67*	0.67*	0.77*	0.66*	0.84 (0.81, 0.87)	0.81	2.24
10. Communication (0–18)	0.77*	−0.60*	−0.50*	−0.55*	0.63*	0.70*	0.55*	0.84 (0.81, 0.87)	0.62	1.71
11. Moving outdoors (0–39)	0.91*	−0.87*	−0.56*	−0.68*	0.68*	0.80*	0.69*	0.89 (0.86, 0.91)	0.93	2.57
12. Gardening and house maintenance (0–21)	0.74*	−0.67*	−0.45*	−0.54*	0.56*	0.67*	0.55*	0.70 (0.65, 0.75)	2.80	7.74
13. Caring (0–27: $n = 275$)	0.32*	−0.29*	−0.25*	−0.29*	0.28*	0.32*	0.25*	0.62 (0.55, 0.67)	3.09	8.54
14. Leisure and social activities(0–27)	0.72*	−0.61*	−0.48*	−0.55*	0.55*	0.68*	0.55*	0.70 (0.64, 0.75)	1.97	5.44

Concurrent validity with comparator measures (Spearman correlations); test–retest reliability (ICC2, 1) and sensitivity to change standard error of measurement (SEM), minimal detectable $change_{95}$ (MDC_{95}) of Rasch-transformed Section A scores. *$P \leqslant 0.001$. Negative correlations occur with SF36v2 scores, as higher scores represent better function in the SF36v2.

Acceptability

At Test 2, 246/294 (83%) of respondents considered the EDAQ helpful or very helpful for discussing everyday problems with an occupational therapist, and 144/165 (87%) considered that it included about the right number of activities.

Discussion

The English EDAQ is a detailed measure of self-reported ability in daily activities developed for use in arthritis rehabilitation. It includes 14 domains, 12 of which can be combined into two components (Self care and Mobility). The EDAQ is unusual as it distinguishes between intrinsic disability (i.e. without ergonomic and environmental modifications) and actual disability (i.e. with such modifications). The activities included are those that people with RA consider most relevant, thus meeting the recommendations for PROMs of the US Food and Drug Administration (FDA) [8].

Part 1 had moderate test–retest reliability. Given that RA symptoms fluctuate over time, this was acceptable over

the average 6-week test–retest period. Both parts have good concurrent validity with other measures of symptoms and PF, and good discriminant validity because domain scores differed significantly with disease status. It also satisfied Rasch model expectations in that Part 2 domains are unidimensional, largely invariant for age, gender and employment status, have high internal consistency and good reliability (apart from the Caring domain, which was moderate, probably because low numbers answered this domain). Further testing of the Caring domain's reliability and validity is needed with a larger number of people with younger children. Sensitivity to change was lower for the three new domains, probably because more respondents reported items as inapplicable: Gardening and household maintenance included activities performed more commonly by men; few items in the Caring domain were widely relevant; while the Leisure and social activities domain reflected personal interests rather than universally essential activities. Scores can be transformed into interval data using a Rasch transformation table [24], allowing calculation of change scores and easier comparison between domains using the MDC_{95}. Each domain and component has robust reliability and validity, meaning either the whole EDAQ or selected domains can be used to identify activity capabilities and problems. Reassessment could include only those domains causing difficulties.

We also established a scoring method to create total scores for each domain and a layout that minimized missing scores. We recommend using the modified layout, particularly if the EDAQ is mailed to participants or there is little opportunity for verbal instructions. EDAQ scoring is now standardized. Section A domain scores cannot be calculated at the present time if there are missing data. Assessors should therefore check the EDAQ and request missing scores and, for research, use multiple imputation. However, in future, developing the domains as item banks (from which tailored tests could be undertaken) could potentially reduce respondent burden and missing value count.

The EDAQ takes most people 20–30 min to complete. This is normally done at home to allow time for reflection on abilities. In this study, we found participants could complete the EDAQ on their own, and most found it acceptable and helpful. The clinical benefits of the EDAQ are that it is comprehensive, and that completing it at home before or between therapy appointments minimizes therapists' assessment time, maximizing time for identifying solutions to problems. This could increase quality and efficiency of services and patient satisfaction.

Similar to many RA rehabilitation and outcome measures studies, a fifth did not return questionnaires. A limitation is that we do not know if non-response was affected by questionnaire length. Given that 16% of the UK population are functionally illiterate, the EDAQ is also not applicable for all patients [38]. We have not tested validity and reliability of the EDAQ administered via interview for those with insufficient reading skills or dislike of completing long questionnaires. We also have not tested responsiveness (for participant-reported health status change or longitudinal construct validity).

The robustness of any psychometric testing depends on generalizability of results. To ensure representativeness, we included 15 UK centres across England and over 500 patients in development, modification and testing. Participants' demographic and disease characteristics were representative of patients referred to rehabilitation, although a higher number than is usual were on biologic medication. There is no generally accepted method of calculating sample size for psychometrics, but we used well-accepted methods for Rasch analysis. Additionally, testing the modified layout meant recruiting a relatively large sample for a psychometric study. Results are thus likely to be generalizable.

In conclusion, the English EDAQ is a comprehensive, psychometrically robust measure of daily activities for people with RA. Future research will determine its applicability for other musculoskeletal conditions. It is suitable for use in research and audit, that is, for evaluating group changes in activity ability. It can also be used clinically to identify clients' problems and aid identification of solutions, but not to evaluate an individual's change in activity ability, because Rasch-transformed ICC (2,1) values were <0.9. Further work is needed to assess responsiveness and the impact of implementation into clinical practice.

Acknowledgements

The authors wish to thank all the study participants; Rachel Gill, Rachel Shuttleworth and Rob Peet, Centre for Health Sciences, University of Salford, for assistance with data collection and data entry; Vikki Lane, Academic Rehabilitation, University of Leeds for assistance with EDAQ layout; Dr Lynne Goodacre, North West Research Design Service for study advice; and all the Principal Investigators, Rheumatology Consultants, rheumatology and research nurses and occupational therapists assisting with participant recruitment and study support at the participating sites: Dr Chris Deighton, Lindsey Kimber, Nicola Watson, Karen Pearson, Alison Booth (Royal Derby Hospital); Prof D.G.I. Scott, Margaret Somerville, Karen Mills, Holly Monton (Norfolk and Norwich Hospital); Dr Mike Green, Esther Gordon, Beverley Neville, Alison Miernik, Anne Gill (Harrogate and District Hospital, York Hospital); Dr Kirstin Mackay, Chris Dixon, Hannah Ruelens, Elaine Wren, Usha Chandra, Patricia Pickford (Torbay Hospital); Prof Ian Bruce, Lindsey Barnes, Dmitrij Bartosi (Manchester Royal Infirmary); Dr Neil Snowden, Denise McSorland (North Manchester General Hospital); Prof David Walsh, Debbie Wilson (King's Mill Hospital, Sutton-in-Ashfield); Dr David James, Heather Gallagher (Princess of Wales Hospital, Grimsby); Emma Jones, Dr Angela Green (Hull Royal Infirmary); Tracey James, Alison Page (Derriford Hospital, Plymouth); Darren Peters, Steve Harvey, Debbie Fuller (Royal Devon and Exeter Hospital); Louise Hollister, Sarah Jervis, Dawn Simmons (Weston Hospital, Weston-super-Mare); Prof A. Jones,

Anne Lenton (Salford Royal Hospital); Dr Kevin Fairburn, Heather Cripps (Chesterfield Royal Hospital).

Funding: This work was supported by Arthritis Research UK [project grant no:18497].

Disclosure statement: The authors have declared no conflicts of interest.

Supplementary data

Supplementary data are available at *Rheumatology* Online.

References

1 Kirwan JR, Reeback JS. Stanford Health Assessment Questionnaire modified to assess disability in British patients with rheumatoid arthritis. Br J Rheumatol 1986; 25:26–9.

2 Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. Arthritis Rheum 1992;35:1–10.

3 Ware JE, Sherbourne CD. The MOS 36-item Short-Form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.

4 Ware JE. SF-36 health survey update. Spine 2000;25:3130–9.

5 White DK, Wilson JC, Keysor JJ. Measures of adult general functional status. Arthritis Care Res 2011;63: S297–307.

6 Gignac M, Cao X, McAlpine J, Badley EM. Measures of disability. Arthritis Care Res 2011;63:S308–24.

7 Kirwan JR, Hewlett SE, Heiberg T *et al*. Incorporating the patient perspective into outcome assessment in rheumatoid arthritis – progress at OMERACT 7. J Rheumatol 2005;32:2250–6.

8 US Department of Health and Human Services Food and Drug Administration. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims, 2009. http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf (15 February 2014, date last accessed).

9 Nordenskiöld U, Grimby G, Hedberg M, Wright B, Linacre JM. The structure of an instrument for assessing the effect of assistive devices and altered working methods in women with rheumatoid arthritis. Arthritis Care Res 1996;9:21–30.

10 Nordenskiold U, Grimby G, Dahlin-Ivanoff S. Questionnaire to evaluate the effects of assistive devices and altered working methods in women with rheumatoid arthritis. Clin Rheumatol 1998;17:6–16.

11 Nordenskiold U. EDAQ Manual Praktisk Genomrade Version 5. Göteborg, Sweden: Institute of Neuroscience and Physiology, Rehabilitation Medicine, Sahlgrenska Academy, Göteborg University, 2007.

12 Cederlund R, Nordenskiold U, Lundborg G. Hand-arm vibration exposure influences performance of daily activities. Disabil Rehabil 2001;23:570–7.

13 Cederlund R, Iwarsson S, Lundborg G. Quality of life in Swedish workers exposed to hand-arm vibration syndrome. Occup Ther Int 2007;14:156–69.

14 Sandqvist G, Eklund MA, Akesson A, Nordenskiold U. ADL activities and hand function in women with scleroderma. Scand J Rheumatol 2004;33:102–7.

15 Thyberg I, Hass UAM, Nordenskiold U, Gerdel B, Skogh T. Activity limitation in rheumatoid arthritis correlates with reduced grip force regardless of sex: the Swedish TIRA Project. Arthritis Care Res 2005;53:886–96.

16 Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 2000;25:3186–91.

17 Hammond A, Tyson S, Prior Y *et al*. Linguistic validation and cultural adaptation of an English version of the Evaluation of Daily Activity Questionnaire in rheumatoid arthritis. Health Qual Life Outcomes 2014;12:143.

18 Linacre JM. Sample size and item calibration stability. Rasch Meas Transact 1994;7:328.

19 Quality Metric Incorporated Solutions, QualityMetric Health Outcomes Scoring Software 4.5. Lincoln, RI: QualityMetric Incorporated 2010. https://www.optum.com/optum-outcomes/what-we-do/health-surveys/sf-36v2-health-survey.html (26 February 2015, date last accessed).

20 Tennant A, Hillman M, Fear J, Pickering A, Chamberlain MA. Are we making the most of the Stanford Health Assessment Questionnaire? Br J Rheumatol 1996;35:574–8.

21 Wolfe F. Which HAQ is best? A comparison of the HAQ, MHAQ and RA-HAQ, a difficult 8 item HAQ (DHAQ), and a rescored 20 item HAQ (HAQ20): analyses in 2491 Rheumatoid Arthritis patients following leflunomide initiation. J Rheumatol 2001;28:982–9.

22 De Jong Z, van der Heijde D, McKenna SP, Whalley D. The reliability and construct validity of the RAQoL: a rheumatoid arthritis-specific quality of life instrument. Br J Rheumatol 1997;36:878–83.

23 Hammond A, Tennant A, Tyson S, Nordenskiold U. The Evaluation of Daily Activity Questionnaire: v1 (parts 1 and 2) (English). Manchester, UK University of Salford, 2014. http://usir.salford.ac.uk/30755/ (11 January 2014, date last accessed).

24 Hammond A, Tennant A, Tyson S, Nordenskiold U. . The Evaluation of Daily Activity Questionnaire: User Manual v1 (English). Manchester, UK: University of Salford, 2014. http://usir.salford.ac.uk/30752/ (11 January 2014, date last accessed).

25 Panter A, Swygert T, Kimberly A, Grant Dahlstrom W, Tanaka JS. Factor analytic approaches to personality item-level data. J Pers Assess 1997;68:561–89.

26 Brown TA. Confirmatory Factor Analysis for applied research. New York: Guilford Press, 2006.

27 Smith EV. Detecting and evaluation the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002;3:205–31.

28 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 Res 2007;57:1358–62.

29 Van Newby A, Conner GR, Bunderson CV. The Rasch model and additive conjoint measurement. J Appl Meas 2009;10:348–54.

30 Masters G. A Rasch model for partial credit scoring. Psychometrika 1982;47:149–74.

31 Andrich D, Sheridan BED, Luo G. RUMM2030: Rasch unidimensional models for measurement. Perth, Western Australia: RUMM Laboratory, 2009.

32 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33: 159–74.

33 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;46:1–18.

34 Stratford PW. Getting more from the literature: estimating the standard error of measurement from reliability studies. Physiother Can 2004;56:27–30.

35 Donoghue D, PROP group and Stokes EK. How much change is true change? The minimum detectable change of the Berg Balance Scale in elderly people. J Rehabil Med 2009;41:343–6.

36 Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998;2:i–iv, 1–74.

37 Terwee CB, Bot SDM, de Boer MR *et al*. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 60:34–42.

38 National Literacy Trust. Words for Life. How many illiterate adults are there in England?. http://www.literacytrust.org.uk/adult_literacy/illiterate_adults_in_england (20 January 2014, date last accessed).