ORIGINAL REPORT

MEASURING MOBILITY IN PEOPLE WITH LOWER LIMB AMPUTATION: RASCH ANALYSIS OF THE MOBILITY SECTION OF THE PROSTHESIS EVALUATION QUESTIONNAIRE

Franco Franchignoni, MD[1], Andrea Giordano, PhD[2], Giorgio Ferriero, MD[1], Duccio Orlandini, MD[3], Amedeo Amoresano, MD[3] and Laura Perucca, MD[4]

From the [1]Unit of Occupational Rehabilitation and Ergonomics and [2]Bioengineering Service, Rehabilitation Institute of Veruno, Salvatore Maugeri Foundation – Clinica del Lavoro e della Riabilitazione, IRCCS, Veruno, [3]Centro Protesi INAIL, Vigorso di Budrio and [4]Unità Clinica di Riabilitazione Neuromotoria, Istituto Auxologico Italiano, IRCCS, Milano, Italy

Objective: To assess the psychometric properties of the Mobility Section of the Prosthesis Evaluation Questionnaire (PEQ-MS).
Design: A postal survey, including self-report assessment of prosthetic capability and performance with the PEQ-MS and the Locomotor Capabilities Index, and of other variables associated with prosthetic wear and use. The PEQ-MS data underwent Rasch analysis for rating scale diagnostics and a reliability and validity study.
Patients: A total of 123 subjects (mean age 54 years) who had undergone lower limb amputation in the previous 5 years and who had completed rehabilitation and a prosthetic training programme.
Results: According to Rasch analysis and expert review, some response categories of the PEQ-MS (13 items, 11-level numeric rating scale) were collapsed and one item was deleted. The remaining 12 items fitted to the Rasch model and created a revised scale with a 5-level response format, the PEQ-MS12/5. The PEQ-MS12/5 demonstrated good reliability (person-separation reliability = 0.95, item-separation reliability = 0.98) and internal construct validity. Moreover, the correlation with the Locomotor Capabilities Index ($r_s = 0.78$) and with prosthetic wear and use (r_s range 0.41–0.59) supported the convergent validity of the PEQ-MS12/5.
Conclusion: The new PEQ-MS12/5 presents good psychometric characteristics for measuring mobility in people with lower limb amputations. These preliminary results provide an already applicable instrument and a solid basis for further validation studies.

Key words: amputation, leg prosthesis, outcome assessment, measurement, Rasch analysis.

J Rehabil Med 2007; 39: 138–144

INTRODUCTION

The best possible restoration of mobility and locomotor function represents the cornerstone of rehabilitation programmes following a lower limb amputation (LLA) (1). Thus, there is increasing interest in psychometrically sound outcome measures of mobility in people with LLA, in order accurately to monitor the impact of therapeutic interventions, in particular of prosthetic trials.

Different approaches have been applied to assess prosthetic mobility (2). They include performance tests (such as the 2-minute walk test (3), the "Timed Up and Go" test (4), and the L Test of Functional Mobility (5)), functional categorizations of ambulation (usually based upon the amount of personal assistance needed and type of aids used) (6), and multi-item ordinal scales (relying on patient reports or observations/examinations by health professionals) (7, 8). There are strengths and limitations associated with each approach (2, 7). One advantage of self-report measures and questionnaires in subjects with LLA is that they allow postal follow-up of long-term results after hospital discharge (9–11).

Among self-report scales investigating locomotor capabilities of people with LLA wearing prostheses, 3 measures have lately received attention in the literature: the Houghton scale (8), the Locomotor Capabilities Index (7, 12, 13) and the Mobility Section of the Prosthesis Evaluation Questionnaire (PEQ-MS) (7, 9, 13, 14). A recent study (7) showed that reliability and validity of the 3 scales are acceptable for group level comparison (according to the classical test theory), but the same study proposed refinements of PEQ-MS and recommended further investigation of its psychometric characteristics, in particular of the best response options.

For a more in-depth validation of basic measurement properties (e.g. scaling properties, internal construct validity, and item bias), outcome measures are increasingly being investigated in practical healthcare applications (15, 16) with respect to their fit to Rasch model requirements (17, 18).

The objective of this study was to perform a Rasch analysis of the PEQ-MS in patients with LLA, in order to investigate the quality of the rating categories and the instrument's validity (unidimensionality and internal construct validity) and reliability. This approach aims to provide the rationale for revising and improving the measurement qualities of this outcome measure designed to capture the overall locomotor ability of people with LLA while wearing a prosthesis.

MATERIALS AND METHODS

Subjects

A questionnaire was posted to 144 subjects who had undergone LLA in the previous 5 years and were consecutively treated by the hospital and prosthesis centre of INAIL (National Institute Against Industrial Injuries and Occupational Diseases) in Vigorso di Budrio-Bologna, Italy. The pre-selection criteria were that subjects: (*i*) had to have completed the rehabilitation period and prosthetic training programme at least 6 months before the start of the study; (*ii*) have been using the prosthesis and have been without mental or physical conditions that would restrict the use of assistive devices for walking, when discharged from rehabilitation and during follow-up; (*iii*) had to be 20–80 years old; (*iv*) had to be living in the community; (*v*) had to be able to read and write. If subjects did not return the questionnaire within 4 weeks a second posting took place, and 20 days later a telephone call was made to non-respondents as a last reminder.

One hundred twenty-three persons out of the 144 study invitees (85%) returned the questionnaire, 111 (77%) at first request. This response rate can be considered excellent (11), and in previous research has been demonstrated as sufficient to eliminate significant non-response biases from the research sample (19).

Instruments

The *Mobility Section of the PEQ (PEQ-MS)* is part of the Prosthesis Evaluation Questionnaire (PEQ), a self-administered questionnaire designed for comprehensive evaluation of persons with amputations "with regard to their prostheses and their prosthesis-related quality of life". It comprises 82 questions, subdivided into 10 scales related to 4 content areas: prosthesis function (4 scales), mobility (2 scales), psycho-social aspects (3 scales), and well-being (1 scale) (9, 20). The PEQ-MS consists of 2 scales (ambulation 8 items; transfer 5 items) that can be combined into a single mobility measure (7). It evaluates the perceived potential for mobility using prosthetic devices over the past 4 weeks. A typical question is "Over the past 4 weeks, rate your ability to…". Most questions in the PEQ use a visual analogue scale (VAS) format, scored as a continuous numerical variable measured as the distance in mm from the left-hand end-point of the 100-mm line to the point at which the respondent's mark crosses the line. This VAS format is unfamiliar and not easily understood by many groups of patients (particularly those with limited education) (7, 21), whereas the postal version of a self-report questionnaire needs to be very easy to complete without assistance (11). For these reasons, we converted, as in previous studies (5, 7, 14), the original VAS format to an 11-step 0–10 numeric rating scale, in order to reduce problems in administration and scoring (22). The end-points were labelled "0 = not able, or hardly able at all (less than 5% of ability)" and "10 = no problems or almost fully able (more than 95% of ability)", in an endeavour to reduce end-aversion bias (23). As suggested by both the authors of the questionnaire (9) and Miller et al. (7), a summary score was calculated as the arithmetical mean of scores from all the questions.

To check the construct validity of the Rasch-refined version of the PEQ-MS, patients were also asked to complete some items from the Prosthetic Profile of the Amputee (PPA), a widely used questionnaire developed and validated for follow-up studies in LLA (13, 24–26). The PPA items were:

- # 11 – Locomotor Capabilities Index (LCI). The LCI is a scale that has often been used separately from the general instrument (7, 26). It comprises 14 questions (phrased as "Would you say that you are able to do the following activities with your prosthesis on?") evaluating subjects' capability in performing different postural tasks and locomotor activities. We adopted the version with a 5-point ordinal scale (ranging from "0 = not able" to "4= able to accomplish the activity without aids") because of its better psychometric properties compared with the original version (12). A composite measure representing the global locomotor ability level is obtained by adding the individual scores assigned to each activity (possible maximum score = 56) (7,12,26);

- # 12 – prosthetic wear (defined as the number of hours the prosthesis was worn per week) (25);
- # 14b and 16b – active use of prosthesis, indoors and outdoors respectively (defined as the percentage of ambulatory activities performed with the prosthesis daily: approximately 0%, 25%, 50%, 75% and 100%) (25);
- # 18 – distance walked non-stop (6-level ordinal scale, from "I do not walk" to "not limited") (25).

Other questions selected from the PPA questionnaire (10, 24, 25), about sociodemographic variables (marital status, education level, employment, income status, etc.), associated medical conditions and amputation-related factors – were also recorded.

Statistical analysis

The median and interquartile range (25th–75th percentile) were used as a measure of central tendency and variability of the score distribution, respectively, given the ordinal nature of the measures (27). Internal consistency was examined through Cronbach's coefficient alpha (23).

The matrix of single raw scores of the PEQ-MS for each subject underwent Rasch analysis through the Winsteps software (28). The following topics were analysed:

- Rating scale diagnostics. As a first step, we investigated whether the rating scale of the PEQ-MS was being used in the expected manner. Our *a priori* hypothesis was that more categories exist in the scale than are needed to describe the construct (23, p. 36). The following criteria – suggested by Linacre (29) – were used to judge this parameter: (*i*) at least 10 cases per category; (*ii*) even distribution of category use; (*iii*) monotonic increase in both average measures across rating scale categories and thresholds. Thresholds, sometimes also called step calibrations, are the ability levels at which the response to either of 2 adjacent categories is equally likely; (*iv*) category outfit mean square values less than 2 (see the following paragraph); (*v*) threshold differences higher than 1.4 logit units and lower than 5. Categories were collapsed following specific guidelines, and several patterns of categorization were compared, looking not only at the above indicators of category diagnostics but also at best reliability indices (see below) (30).

- Reliability and validity. Reliability was evaluated in terms of "separation" (G), defined as the ratio of the true spread of the measures with their measurement error (17, 30). The item-separation index gives an estimate (in standard error units) of the spread or "separation" of items along the measurement construct; the person-separation index gives an estimate of the spread or separation of persons along the measurement construct. This index reflects the number of "strata" of measures which are statistically discernible. A separation of 2.0 is considered good and enables the distinction of 3 groups or strata, defined as segments whose centres are separated by distances greater than can be accounted for by measurement error alone [number of distinct strata = (4G + 1)/3] (17). A related index is the reliability of these separation indices, providing the degree of confidence that can be placed in the consistency of the estimates (range 0–1; coefficients > 0.80 are considered good, and > 0.90 excellent) (30).

After the revision of the rating scale categories, validity was analysed by evaluating the fit of individual items to the latent trait as per the Rasch modelling (unidimensionality) and determining whether the pattern of item-difficulties was consistent with the model expectation. For Rasch analysis it is reported that a sample size of about 100 persons will estimate item-difficulty with an alpha of 0.05 to within ± 0.5 logits (31).

Depending on the string of responses provided by a particular sample of subjects on a particular sample of items, the Rasch model estimates goodness-of-fit (or simply "fit") of the real data to the modelled data. Information-weighted (infit) and outlier-sensitive (outfit) mean-square statistics (MnSq) for each item were calculated (similarly to a χ^2 analysis) to test if there were items that did not fit with the model expectancies. In accordance with the literature, we considered as an

indicator of acceptable fit MnSq > 0.6 and < 1.4: items outside this range were considered misfitting (MnSq ≥ 1.4) or overfitting (MnSq ≤ 0.6) (see below) (30).

The next step in the Rasch analysis was to calculate the level of difficulty achieved by each item (item-difficulty) and where each individual subject fits along the continuum (subject ability). Item-difficulty and patient ability are expressed – on a common interval scale – in logit units, a logit being the natural logarithm of the ratio (odds) of mutually exclusive alternatives (e.g. pass vs fail, or higher vs lower response) (17, 30).

Finally, to test the construct validity of the Rasch-refined version of the PEQ-MS, we correlated its score with LCI and other PPA items (Spearman's r_s, corrected for ties) (27), hypothesizing a good to excellent correlation with LCI (the constructs measured by the 2 scales are both related to ability in locomotor activities with the prosthesis) and a fair to moderate degree of relationship with the other PPA items measuring factors associated with prosthetic wear and use (25).

RESULTS

Table I shows the main socio-demographic characteristics of the 123 persons who returned the envelope. The median value (25th–75th percentile) of the PEQ-MS was 6.6 (4.9–8.1) on a 0–10 scale; 5 out of the 123 subjects (4%) showed the PEQ-MS maximum score and nobody the minimum score.

Some levels of the item's rating categories in PEQ-MS did not comply with the pre-set criteria for category functioning (average measures, thresholds, etc.) (Fig. 1A and Table II). For this reason, the 11 original PEQ scale categories were combined in different ways until the criteria were best met. This was obtained using 5 categories and adopting the following rescaling: 0 = 0; 1–3 = 1; 4–6= 2; 7–9 = 3; 10 = 4. The 5 resulting categories roughly represent the following levels: 0 = not able or hardly able at all (less than 5% ability)"; 1 = high difficulty (5–34% ability); 2 = moderate difficulty (35–64% ability); 3 = little difficulty (65–95% ability); 4 = no problems or almost fully able (more than 95% ability). Fig. 1B shows category probability curves after the collapsing procedure.

Rasch analysis showed that 12 of the 13 PEQ-MS items fitted the unidimensional construct that the scale was intended to measure (MnSq between 0.6 and 1.4) (Table III). The item "Shower/bathe" was misfitting (Infit MnSq = 1.59; Outfit Mnsq = 1.55). The revised scale including only the 12 items of the PEQ-MS fitting the Rasch model (all but the item "Shower/bathe") constitutes a new Rasch-refined instrument with an original 5-level response format, the PEQ-MS12/5 (see Appendix).

The reliability indices for the PEQ-MS before and after the phase of rating-scale modification and item reduction are shown in Table IV. The reliability indices of PEQ-MS12/5 presented comparable values with those of the PEQ-MS; all item and person reliabilities were ≥ 0.95. The items of both versions were distributed into more than 10 difficulty strata.
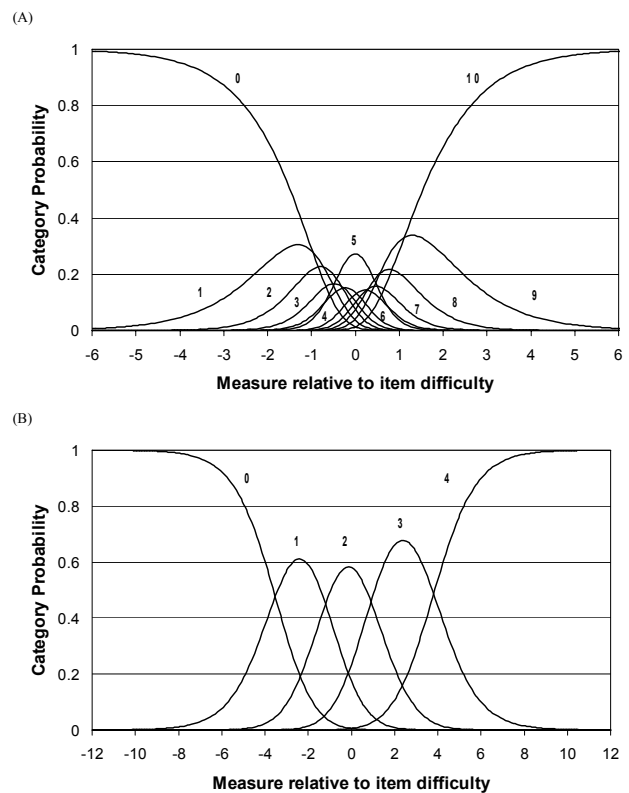
(A)

(B)

Fig. 1. Category probability curves of the Prosthesis Evaluation Questionnaire – Mobility Section (PEQ-MS). (A) Original scale with 11 categories (0–10). (B) Revised scale after collapsing into 5 categories, 0–4 (PEQ-MS12/5). The y-axis represents the probability (0 to 1) of responding to one of the rating categories and the x-axis represents the different performance values (person ability minus the item difficulty), in logits. The "0" curve declines as the subject's ability increases; the crossing point (where 0 and 1 are equally probable) is the first "threshold". The same applies for the other curves. The plot should look – as in Fig. 1B – like a range of hills, each with an "emerging" top. It can be seen from Fig. 1A that the probability of using the central categories is never higher than that of adjacent ratings (except for category 5). Conversely, in Fig. 1B it can be seen that the probability of selecting each of the 5 revised rating categories (0–4) is now a clear function of the level of ability shown by the subject in the x-axis. Correspondingly, the "thresholds" are ordered as intended (e.g. a greater ability is required when the most likely response is 1 rather than 0, 2 rather than 1, etc.).

Table I. *Characteristics of study population (n = 123). Median values (25th–75th percentile, interquartile range (IQR)) or percentages*

Characteristic	Value (IQR)
Age, years (range)	54 (36–65)
Gender, male (%)	74
Educational level, years (range)	8 (5–14)
Time since amputation, months (range)	38 (18–48)
Cause of amputation	
Peripheral vascular disease and/or diabetes mellitus (%)	35
Trauma (%)	56
Tumour and other (%)	9
Amputation level	
Unilateral, above the knee (%)	53
Unilateral, below the knee (%)	36
Bilateral (%)	11 (6% above both knee)
Current vocational status	
Employed (%)	28
Unemployed or retired (%)	72

Table II. *Prosthesis Evaluation Questionnaire – Mobility Section (PEQ-MS) before (left, 11 categories) and after (right, 5 categories) the collapsing procedure (n = 118 persons with non-extreme measures): frequencies, average category measures (mean ability of subjects getting a given score) and thresholds (step calibration, ability level at which adjacent scores are equally likely)*

PEQ-MS with 11 categories					PEQ-MS with 5 categories				
Category label	Observed count	%	Average measure	Step calibration	Category label	Observed count	%	Average measure	Step calibration
0	71	5	−1.41	None	0	71	5	−3.06	None
1	65	4	−1.09	−1.20					
2	67	4	−0.73	−0.93	1	231	15	−1.56	−3.52
3	99	6	−0.33	−1.00					
4	100	7	−0.21	−0.33					
5	161	10	0.02	−0.52	2	397	26	0.21	−1.20
6	136	9	0.46	0.42					
7	122	8	0.61	0.63					
8	162	11	0.93	0.54	3	510	33	2.17	0.91
9	226	15	1.47	0.88					
10	325	21	2.37	1.52	4	325	21	4.70	3.81

Furthermore, PEQ-MS12/5 was able to distinguish more than 5 levels of patient ability (very low, low, average, high, very high) in this study sample.

Regarding the hierarchic ordering of items, Fig. 2 shows the distribution map of subject ability and item-difficulty of PEQ-MS, according to the Rasch model. Subjects' ability levels (average measure = 1.33) spanned more than 10 logits (from −3.85 to +6.94 logits): distribution was symmetric around the mean. The mean error estimate for the person ability levels was 0.48; as expected, in the near-maximum scores (6.63 logits) the standard error (SE) was higher (SE = 1.08, about 16% of the corresponding measure). Item-difficulty estimates spanned 4.32 logits (from −1.97 to +2.35 logits) (see Table III, including SE for each item). The lowest difficulty threshold estimate (step from category 0 to 1) of item 10 "Sit down and get up from a chair with a high seat" (the easiest activity) was near

Table III. *Item calibrations (measure, increasing bottom-up) with standard errors (SE), and infit and outfit mean-square statistics (MnSq) for the 13 items of the Prosthesis Evaluation Questionnaire – Mobility Section (PEQ-MS). Misfitting values are in bold. The difficulty estimate for the item as a whole is the mean value of the difficulty measures of the thresholds along that item. The higher the item estimate, the less likely it is for any subject to get a high score in that item*

Item	Measure	SE	Infit MnSq	Outfit MnSq
8. Walk on slippery surfaces	2.35	0.15	1.24	1.30
6. Walk down a steep hill	1.92	0.15	0.87	0.88
5. Walk up a steep hill	1.44	0.15	0.75	0.76
11. Sit down and get up from a low or soft chair	0.29	0.15	1.16	1.12
13. Shower or bathe	0.24	0.15	**1.59**	**1.55**
4. Walk down stairs	−0.36	0.15	0.82	0.77
2. Walk in close spaces	−0.41	0.16	0.90	0.92
3. Walk up stairs	−0.41	0.16	0.83	0.81
9. Get in and out of a car	−0.54	0.16	0.92	0.85
12. Sit down and get up from the toilet	−0.66	0.16	1.23	1.19
7. Walk on sidewalks and streets	−0.87	0.16	0.91	0.88
1. Walk	−1.11	0.16	0.98	1.01
10. Sit down and get up from a chair with a high seat	−1.97	0.17	0.89	0.82

Table IV. *Reliability indices of Prosthesis Evaluation Questionnaire – Mobility Section (PEQ-MS) (13 items, 11 rating categories) and PEQ-MS12/5 (12 items, 5 rating categories)*

	PEQ-MS	PEQ-MS12/5
Item-separation index	7.48	7.90
Item-separation reliability	0.98	0.98
Person-separation index	4.20	4.15
Person-separation reliability	0.95	0.95
Cronbach's alpha	0.96	0.96

−5 logits, whereas the highest threshold (step from category 3 to 4) of item 8 "Walk on slippery surfaces" (the most difficult activity) was about 6 logits. The ability levels of the study sample were rather well-matched with the difficulty levels of the items, except for those participants above 6 logits, for whom mobility (at least as analysed with PEQ-MS) does not appear to be a problem.

Regarding the construct validity of the PEQ-MS12/5, this new scale strongly correlated with LCI (r_s = 0.78), and showed a moderate but significant correlation with the other PPA items (#12 – frequency of prosthetic wear: r_s = 0.59, #14b – active use of prosthesis indoors: r_s = 0.43, #16b – active use of prosthesis outdoors: r_s = 0.41, and #18 – distance walked non-stop r_s = 0.48; p < 0.001 for all). The original 11-level PEQ-MS presented similar, but slightly lower, correlations (not shown).

DISCUSSION

Previous analyses performed according to the classical test theory demonstrated that the PEQ-MS has high internal consistency and test-retest reliability and shows convergent validity, as it correlated with the 2-minute walking test, the Timed Up and Go Test, the Activity-specific Balance Confidence Scale, and the LCI (7, 13, 14). These preliminary studies recommended further investigation of the psychometric characteristics of the scale. In particular, Miller et al. (7) suggested additional research in this field using a demanding approach such as Rasch analysis. The purpose of our study was to perform a Rasch
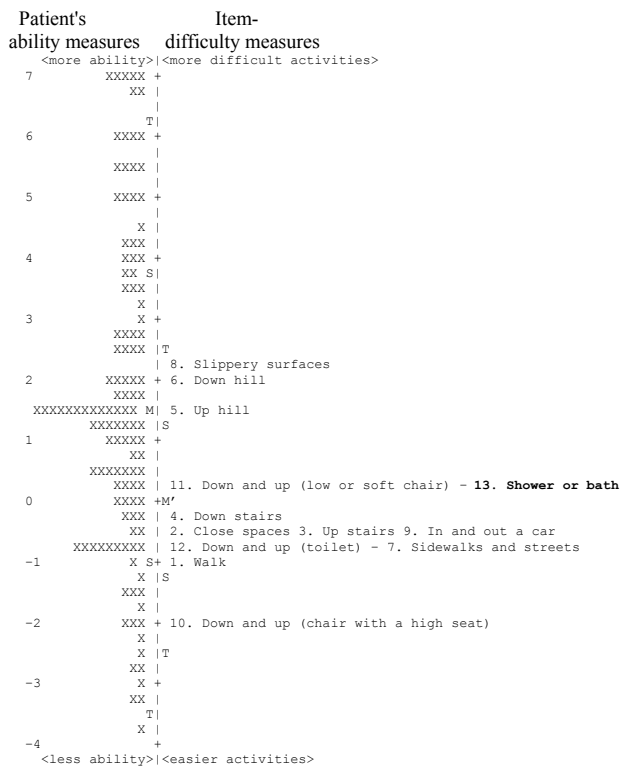
```
Patient's              Item-
ability measures    difficulty measures
      <more ability>|<more difficult activities>
  7        XXXXX +
            XX |
               |
             T|
  6        XXXX +
               |
           XXXX |
               |
  5        XXXX +
               |
             X |
           XXX |
  4        XXX +
          XX S|
           XXX |
             X |
  3          X +
          XXXX |
          XXXX |T
               | 8. Slippery surfaces
  2        XXXXX + 6. Down hill
           XXXX |
  XXXXXXXXXXXXXX M| 5. Up hill
       XXXXXXX |S
  1        XXXXX +
            XX |
          XXXXXXX |
           XXXX | 11. Down and up (low or soft chair) - 13. Shower or bath
  0        XXXX +M'
           XXX | 4. Down stairs
            XX | 2. Close spaces 3. Up stairs 9. In and out a car
  XXXXXXXXXX | 12. Down and up (toilet) - 7. Sidewalks and streets
 -1        X S+ 1. Walk
             X |S
           XXX |
             X |
 -2        XXX + 10. Down and up (chair with a high seat)
             X |
             X |T
            XX |
 -3          X +
            XX |
             T|
             X |
 -4          +
      <less ability>|<easier activities>
```

Fig. 2. Person ability and item-difficulty maps (so-called "Rasch ruler") of the Prosthesis Evaluation Questionnaire – Mobility Section (PEQ-MS). The vertical line represents the measure of the variable, in linear logit units. The left-hand column locates the individual's ability along the variable: each person is indicated by an "X". The right-hand column locates the 13 item-difficulty measures along the variable (for each item, the difficulty estimate represents the mean calibration of the threshold parameters according to the rating scale model). Each item is also indicated by its number in the original questionnaire. Misfitting items are in bold. From bottom to top measures indicate greater mobility (for patients) and greater difficulty (for items), respectively. By convention, the average difficulty of items in the test is set at 0 logits (and indicated with M'). Accordingly, a candidate with average ability is indicated with M. The higher the difference between the patient's and item's measure, the higher the score expected.

analysis of PEQ-MS in patients with LLA in order to improve it by refining the rating scale, distinguishing items belonging to the same construct, verifying the expected difficulty hierarchy of items, and analysing test reliability.

The results demonstrate that the newly designed scale PEQ-MS12/5 show psychometric properties that compare favourably with those of the original PEQ-MS, despite a 1-item reduction and rating scale simplification.

As expected, the 11 categories of PEQ-MS showed some disordered thresholds and had to be collapsed in order to comply with the pre-set quality criteria for rating categories (29). Disordered thresholds suggested that the questionnaire was unable to distinguish subjects' abilities as finely as suggested by the 11 response categories. Proper collapsing usually improves the measure (30), eliminating the redundancy of under-utilized rating categories and ensuring that each rating category represents a distinct level of ability, compared with the adjacent ones (32, 33). As an example, a recent Rasch study showed that a 4-category (instead of 7-category) Functional Independence Measure (FIM™) scale might be even more precise in measuring the level of dependence in the routine clinical setting (32). Similar problems and solutions have also been illustrated recently by Zhu (34).

After collapsing the categories, the data were analysed again in order to extract Rasch-modelled parameters of ability and difficulty, and the validity and reliability issues were faced again.

One item of the PEQ-MS ("Shower/bathe") proved, in fact, to be "misfitting", i.e. poorly related to the measured underlying dimension. The misfit could be due, among other reasons, to its being part of another construct, poorly written, or too sensitive to confounding factors (27, 35). We suggest the elimination of this item for both statistical and content reasons, because this task is sensitive to cultural and environmental factors. Indeed, a similar misfit was observed for the item "Bathing" of the Rivermead Mobility Index in both patients after stroke (36) and LLA (37).

No item was "overfitting" (MnSq < 0.6), i.e. too predictable in its responses, hence little informative. Overall, these findings confirm the general adequacy of the item selection made by the original authors (9), and the correctness of combining the 2 PEQ scales (ambulation and transfer) into a single mobility measure (7). The items provide pertinent clinical information with regard to the actual level of performance of subjects with LLA. The person ability and item-difficulty map for PEQ-MS (Fig. 1) showed a wide logit range for both these variables. The targeting and spread of item-difficulty and the high person-separation reliability showed that the scale is appropriate for measuring with satisfactory precision persons with a wide variety of functional ability. The 4% of subjects with extreme maximum scores – the 5 "X" at the top of the left-hand column in Fig. 2 – constitute a minor trend toward a ceiling effect in very highly functioning subjects (the precision of their ability estimates is lower but still acceptable: the error is about 16% of the corresponding measure). Similarly, the high item-separation reliability indicates that great confidence can be placed also in the replicability of item-placement across future samples. Moreover, the Cronbach's alpha values also indicate that both PEQ-MS and PEQ-MS12/5 have excellent internal consistency: the same value (0.96) with and without item 13 is further evidence that this item is not informative and can be deleted.

As a further demonstration of the construct validity of PEQ-MS12/5 (and also PEQ-MS), the general hierarchic arrangement found by Rasch analysis in the present report is consistent with clinical expectations. For example, "Walk on slippery surfaces" (item 8) was selected as the most demanding task; also, seat height influenced the item-difficulty of rising from a chair (item 11 more difficult than item 10). In addition, the PEQ-MS12/5 showed – as expected – a strong correlation with LCI (due to the close relationship between the

constructs measured by the 2 scales, both related to the ability in locomotor activities with prostheses) and a fair to moderate degree of relationship with the other variables associated with prosthetic use, such as frequency of prosthetic wear, active use of prosthesis, and distance walked non-stop.

In conclusion, Rasch analysis allowed the improvement of the rating scale of PEQ-MS (through category diagnostics) and the identification of those items that are most useful to measurement of the intended construct (as per the indexes of unidimensionality and internal construct validity), and it showed that one can place confidence in the consistency (reliability) of both person ability and item-difficulty estimates obtained by the PEQ-MS12/5. Further studies are needed. They should include: (*i*) analysis of the actual performance of the new response structure, because collapsing 11 categories to 5 is not the same as presenting 5 categories to respondents; (*ii*) a study of differential item functioning, i.e. the stability of item hierarchy across sub-samples defined according to potentially relevant clinical criteria (e.g. age, type of prosthesis, etc.) (38).

REFERENCES

1. Geertzen JH, Martina JD, Rietman HS. Lower limb amputation. Part 2: rehabilitation – a 10 year literature review. Prosthet Orthot Int 2001; 25: 14–20.
2. Rommers GM, Vos LD, Groothoff JW, Eisma WH. Mobility of people with lower limb amputations: scales and questionnaires: a review. Clin Rehabil 2001; 15: 92–102.
3. Brooks D, Hunter JP, Parsons J, Livsey E, Quirt J, Devlin M. Reliability of the two-minute walk test in individuals with transtibial amputation. Arch Phys Med Rehabil 2002; 83: 1562–1565.
4. Schoppen T, Boonstra A, Groothoff J, de Vries J, Göeken LNH, Eisma WH. The Timed "up and go" test: reliability and validity in persons with unilateral lower limb amputation. Arch Phys Med Rehabil 1999; 80: 825–828.
5. Deathe AB, Miller WC. The L test of functional mobility: measurement properties of a modified version of the timed "up & go" test designed for people with lower-limb amputations. Phys Ther 2005; 85: 626–635.
6. Gailey RS, Roach KE, Applegate EB, Cho B, Cunniffe B, Licht S, et al. The Amputee Mobility Predictor: an instrument to assess determinants of the lower-limb amputee's ability to ambulate. Arch Phys Med Rehabil 2002; 83: 613–627.
7. Miller WC, Deathe AB, Speechley M. Lower extremity prosthetic mobility: a comparison of 3 self-report scales. Arch Phys Med Rehabil 2001; 82: 1432–1440.
8. Devlin M, Pauley T, Head K, Garfinkel S. Houghton Scale of prosthetic use in people with lower-extremity amputations: reliability, validity, and responsiveness to change. Arch Phys Med Rehabil 2004; 85: 1339–1344.
9. Legro MW, Reiber GD, Smith DG, del Aguila M, Larsen J, Boone D. Prosthesis Evaluation Questionnaire for persons with lower limb amputation; assessing prosthesis-related quality of life. Arch Phys Med Rehabil 1998; 79: 931–938.
10. Gauthier-Gagnon C, Grise MC. Prosthetic profile of the amputee questionnaire: validity and reliability. Arch Phys Med Rehabil 1994; 75: 1309–1314.
11. McColl E, Jacoby A, Thomas L, Soutter J, Bamford C, Steen N, et al. Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients. Health Technol Assess 2001; 5: 1–256.
12. Franchignoni F, Orlandini D, Ferriero G, Moscato TA. Reliability, validity, and responsiveness of the Locomotor Capabilities Index in adults with lower-limb amputation undergoing prosthetic training. Arch Phys Med Rehabil 2004; 85: 743–748.
13. Ferriero G, Dughi D, Orlandini D, Moscato T, Nocita D, Franchignoni F. Measuring long-term outcome in people with lower limb amputation: cross-validation of the Italian versions of the Prosthetic Profile of the Amputee and Prothesis Evaluation Questionnaire. Eura Medicophys 2005; 41: 1–6.
14. Miller WC, Deathe AB, Speechley M, Koval J. The influence of falling, fear of falling, and balance confidence on prosthetic mobility and social activity among individuals with a lower extremity amputation. Arch Phys Med Rehabil 2001; 82: 1238–1244.
15. Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation. J Rehabil Med 2003; 35: 105–115.
16. Conrad KJ, Smith EV. International conference on objective measurement: applications of Rasch analysis in health care. Med Care 2004; 42 Suppl 1: I1–6.
17. Wright BD, Masters GN, editors. Rating scale analysis. Chicago, IL: MESA Press; 1982.
18. Wright BD, Stone MH, editors. Best test design. Chicago, IL: MESA Press; 1979.
19. Siemiatycki J, Campbell S. Nonresponse bias and early versus all responders in mail and telephone surveys. Am J Epidemiol 1984; 120: 291–301.
20. Prosthetics Research Study. Prosthesis Evaluation Questionnaire [cited 2006 Jul 5]. Available from: http://www.prs-research.org/htmPages/PEQ.html.
21. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998; 2: 1–74.
22. Jaeschke R, Singer J, Guyatt GH. A comparison of seven-point and visual analogue scales. Data from a randomized trial. Control Clin Trials 1990; 11: 43–51.
23. Streiner DL, Norman GR, editors. Health measurement scales. A practical guide to their development and use. 2nd edn. Oxford: Oxford University Press; 1995.
24. Grisé MCL, Gauthier-Gagnon C, Martineau GG. Prosthetic profile of people with lower extremity amputation: conception and design of a follow-up questionnaire. Arch Phys Med Rehabil 1993; 74: 862–870.
25. Gauthier-Gagnon C, Grisé MC, Potvin D. Enabling factors related to prosthetic use by people with transtibial and transfemoral amputation. Arch Phys Med Rehabil 1999; 80: 706–713.
26. Gauthier-Gagnon C, Grisé MC, Lepage Y. The Locomotor Capabilities Index: content validity. J Rehabil Outcomes Meas 1998; 2: 40–46.
27. Portney LG, Watkins MP, editors. Foundations of clinical research: applications to practice. 2nd ed. Upper Saddle River: Prentice-Hall Health; 2000.
28. Linacre JM. A user's guide to WINSTEPS Rasch-model computer programs [cited 2006 Jul 5]. Available from: http://www.winsteps.com/aftp/winsteps.pdf.
29. Linacre JM. Investigating rating scale category utility. J Outcome Meas 1999; 3: 103–122.
30. Bond TG, Fox CM, editors. Applying the Rasch model: fundamental measurement in the human sciences. Mahwah, NJ: Lawrence Erlbaum Associates; 2001.
31. Linacre JM. Sample size and item calibration stability. Rasch Meas Trans 1994; 7: 328.
32. Nilsson AL, Sunnerhagen KS, Grimby G. Scoring alternatives for FIM in neurological disorders applying Rasch analysis. Acta Neurol Scand 2005; 111: 264–273.
33. Kornetti DL, Fritz SL, Chiu YP, Light KE, Velozo CA. Rating scale analysis of the Berg Balance Scale. Arch Phys Med Rehabil 2004; 85: 1128–1135.
34. Zhu W. A confirmatory study of Rasch-based optimal categorization of a rating scale. J Appl Meas 2002; 3: 1–15.
35. McNamara TF, editors. Measuring second language performance.

Harlow: Addison Wesley Longman; 1996, p. 149–181.

36. Franchignoni F, Tesio L, Benevolo E, Ottonello M. Psychometric properties of the Rivermead Mobility Index in Italian stroke rehabilitation inpatients. Clin Rehabil 2003; 17: 273–282.

37. Ryall NH, Eyres SB, Neumann VC, Bhakta BB, Tennant A. Is the Rivermead Mobility Index appropriate to measure mobility in lower limb amputees? Disabil Rehabil 2003; 25: 143–145.

38. Tennant A, Penta M, Tesio L, Grimby G, Thonnard JL, Slade A, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care 2004; 42 Suppl 1: I37–48.

APPENDIX

The revised Prosthesis Evaluation Questionnaire – Mobility Section, PEQ-MS12/5, the newly proposed scale for measuring mobility in people with lower limb amputation (LLA) wearing a prosthesis. The scale includes the 12 items of the PEQ-MS fitting the Rasch model, rated with a 5-level scale.

Over the past 4 weeks, please rate your ability in the following activities when using your prosthesis:
"Check ☒ for each statement"

	Unable or hardly able at all (ability < 5%) (0)	High difficulty (ability 5–34%) (1)	Moderate difficulty (ability 35–64%) (2)	Little difficulty (ability 65–95%) (3)	No problems or almost fully able (ability > 95%) (4)
1. To walk	☐	☐	☐	☐	☐
2. To walk in confined spaces	☐	☐	☐	☐	☐
3. To walk upstairs	☐	☐	☐	☐	☐
4. To walk downstairs	☐	☐	☐	☐	☐
5. To walk up a steep hill	☐	☐	☐	☐	☐
6. To walk down a steep hill	☐	☐	☐	☐	☐
7. To walk on sidewalks and streets	☐	☐	☐	☐	☐
8. To walk on slippery surfaces (e.g. wet tile, snow, a rainy street, or a boat deck)	☐	☐	☐	☐	☐
9. To get in and out of a car	☐	☐	☐	☐	☐
10. To sit down and get up from a chair with a high seat (e.g. a dining chair, an office chair)	☐	☐	☐	☐	☐
11. To sit down and get up from a low, soft chair (e.g. a deep sofa)	☐	☐	☐	☐	☐
12. To sit down and get up from the toilet of regular height (no aids)	☐	☐	☐	☐	☐