ABILHAND: A Rasch-Built Measure of Manual Ability

Massimo Penta, Ing, Jean-Louis Thonnard, PhD, Luigi Tesio, MD

ABSTRACT. Penta M, Thonnard J-L, Tesio L. ABILHAND: a Rasch-built measure of manual ability. Arch Phys Med Rehabil 1998;79:1038-42.

Objective: To apply the Rasch measurement model to the development of a clinical tool for measuring manual (dis)ability (ABILHAND).

Design: Manual ability was evaluated in terms of the difficulty perceived by a hand-impaired patient on 57 representative unimanual or bimanual activities.

Setting: A clinical laboratory.

Patients: Eighteen rheumatoid arthritis patients (14 women, 4 men) were interviewed after wrist arthrodesis (10 right, 4 left, and 4 both wrists). Their ages ranged from 38 to 77 years, time since diagnosis ranged from 7 to 41 years, and time since surgery ranged from 0.5 to 17 years.

Main Outcome Measure: ABILHAND, administered at a mean duration of 7 years after arthrodesis.

Results: Forty-six of the 57 items define a common, single manual ability continuum with widespread measurement range and regular item distribution. Items relating to feeding, grooming, and dressing upper body worked consistently with their counterparts in other disability scales. More difficult items extend the measurement range beyond that of most existing manual ability scales.

Conclusion: Even in a small sample of patients, using the Rasch methodology enabled the investigators to produce a useful scale of manual (dis)ability and to define manual ability as a unique construct, at least in patients with rheumatoid arthritis.

FUNCTIONAL INTEGRITY of the hand and upper limb is essential for many activities, yet it is lost as a result of a wide variety of impairments. Therefore, many tests have been developed to assess either the functional losses of mobility, grip strength, tactile sensitivity, and dexterity,[1,2] or a patient's ability to execute manual activities.[3-11] The relation between impairment (dysfunction at the organ/segment level) and disability (dysfunction at the person level)[12] is not straightforward. A patient may adopt either intrinsic or adaptative recovery mechanisms,[13] the latter depending on the integrity of the unaffected organs/segments and on a complex interaction between psychosocial (eg, motivation), cognitive (eg, memory, attention, and space perception), and sensorimotor skills. Thus, disability must be measured per se and not merely inferred from the underlying impairments. Ability can be measured as the capacity of a person as a whole to execute activities, scaled along a continuous gradient of difficulty. But the question remains: which activities should be selected, and how should they be scored?

For standardization purposes, many existing manual ability tests are built on artificial and/or mainly unimanual tasks.[4,7,9] Moreover, the scoring criterion is generally restricted to the observation of a patient's performance in different activities within an artificial, highly motivating clinical environment. These considerations limit a test's adequacy to evaluate correctly a patient's performance in actual daily life activities. Other limitations arise from tests developed for specific levels of performance[6] but lacking sensitivity at higher levels (eg, in the later stages of recovery). Also, in most existing manual ability tests, raw ordinal scores are misused as measures,[3,4,6,8,9,11] when in fact they are merely ranks unsuited to conventional arithmetic.[14,15] Likewise, timed tests[5,7,10] are only ordinal in nature, insofar as a patient being twice as fast as another is not necessarily twice as able in performing the task.[16]

The purpose of this study was to build a valid scale of manual ability. We selected a sample of patients with wrist arthrodesis, after rheumatoid arthritis (RA). In fact, these patients exhibited homogeneous hand and upper limb impairments in the absence of more general movement-disturbing impairments (eg, spasticity, paresis, or ataxia). The Rasch model[17,18] appeared to be the ideal framework to achieve a valid measurement scale.[15] The test was constructed as a self-estimation report to capture a patient's average feeling of difficulty in natural tasks.[19] The selected items are centered on high ability levels where patients are functionally independent but still differ in their performance of manual activities. Therefore, the performance criterion, not the dependence criterion, was adopted for scoring.[20]

METHODS

Sample

The subjects were selected from among the 33 patients with RA who underwent arthrodesis of one or both wrists in the past 25 years at a university clinic. The same surgeon performed the surgery on all the patients. At the time of evaluation, one patient had died, six had changed their address and could not be located, and eight had difficulty moving and could not come to the laboratory. The 18 remaining patients (4 men and 14 women) took part in the study. Their ages ranged from 38 to 77 years (mean 59). The patients were evaluated between 6.8 and 40.9 years (mean 16.9) after diagnosis of RA. At the time of evaluation, none of the patients had an acute inflammatory disease. All patients were right-handed; 10 had undergone the arthrodesis on the right wrist, 4 on the left wrist, and 4 on both wrists. The surgical procedure, derived from the Mannerfelt and Malmsten operation for wrist fusion,[21] fixed the wrist in slight extension and ulnar deviation while maintaining pronation and supination of the forearm. The patients were tested between 0.5 and 16.8 years (mean 7.1) after the arthrodesis.

From Rehabilitation and Physical Medicine Unit, Catholic University of Louvain, Brussels, Belgium (Mr. Penta, Dr. Thonnard); and the Department of Rehabilitation and Department of Research, Functional Assessment and Quality Assurance, "Salvatore Maugeri" Foundation, Instituto di Ricoverdo e Cura a Carattere Scientifico, Pavia, Italy (Dr. Tesio).

Submitted for publication January 19, 1998. Accepted February 20, 1998.

Supported by a grant from the Belgian Association Nationale d'Aide aux personnes Handicapées.

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the authors or upon any organization with which the authors are associated.

Reprint requests to J.L. Thonnard, PhD, Rehabilitation and Physical Medicine Unit, Catholic University of Louvain, Tour Pasteur (5375), Avenue Mounier 53, 1200 Brussels, Belgium.

0003-9993/98/7909-4825$3.00/0

Instrument

The 57 items comprising the ABILHAND questionnaire include both unimanual and bimanual activities, which explore the most representative inventory of manual activities. Some items were selected from existing scales; others were devised to extend the range of activities explored by the questionnaire. To obtain the greatest resolution in measuring manual ability, we selected only items for which manual skills are vital for successful performance. Patients were evaluated with the French version of the questionnaire.

Procedures

The test was administered as a self-reported anamnestic questionnaire. Patients were asked to estimate their ease or difficulty in performing each activity, without technical or human help, on a six-level rating scale where 0 = impossible, 1 = very difficult, 2 = difficult, 3 = fairly easy, 4 = easy, and 5 = very easy. Patients were requested not to rate activities they had never performed (these scores were recorded as missing in the analysis). Patients were asked to use the whole scale range to respond, and five training items were presented before the actual test. The 57 items were then randomly presented via a computer program, and each item was read by the examiner to the patient. For both unilateral and bilateral activities, patients were asked to give their response irrespective of the limb they would actually use to perform the task. The response was then entered into the computer by the examiner.

During the evaluations, patients reported some difficulty in discriminating six response levels when answering the items. This was confirmed by the very low probability for patients to respond in the categories "very difficult" and "fairly easy." According to Linacre[22] and Andrich,[23] this indicates that a threshold between the categories "impossible" and "very difficult" was not clearly identified; neither was it identified between the categories "difficult" and "fairly easy." This suggests that the six-level rating scale offered more categories than the patients were actually able to distinguish. Therefore, the response categories "impossible" and "very difficult" were combined as "very difficult"; and the categories "difficult" and "fairly easy" as "slightly difficult." Hence, the four response categories maintained became 0 = very difficult, 1 = slightly difficult, 2 = easy, and 3 = very easy. The 57 items were analyzed according to this response scale.

Data Analysis

The Rasch model: from raw scores to measures. The data were analyzed with the BIGSTEPS Rasch analysis computer program,[24] according to a rating scale model.[18] The Rasch model[17] is a prescriptive, rather than descriptive model. It requires that item difficulty and patient ability, only, determine the probabilities of category choices when the patient and item interact. From the matrix of response raw scores, the model estimates a linear ability for each patient and a linear difficulty for each item (item calibration). These measures are scaled along a unidimensional continuum delineating the variable purported to be measured by the items, and ranging from minus to plus infinity. The model provides measures that are independent of the particular sample of patients and items, giving rise to a sample-free scale calibration and test-free measures. Measurement units are usually expressed in logits, the logarithm of the ratio of "pass" and "fail" probabilities. Conventionally, a zero measure is set at the average item difficulty. Raw scores represent discontinuous levels, with unknown distances in between. Furthermore, they cannot represent the same quantity along the whole scale.[14,15] Unavoidably, much wider levels of ability are compressed around the extremes of the finite raw score scale. By contrast, the logit metric provides a true interval (linear) unit, ie, it represents a fixed increment along the whole scale of the explored variable. The Rasch model can analyze dichotomous (yes/no) scores as well as polytomous (eg, no/mild/moderate/severe) scores. These analyses do not require complete data. Hence, missing data are not a problem.

Unidimensionality: the analysis of "fit." Patient measures and item calibrations are accompanied by standard errors describing the range within which true patient abilities and item difficulties are expected to lie. The analysis also provides fit statistics indicating the extent to which the data meet the model specification of unidimensionality. If items align along a single continuum, one would expect patients of any ability level to have better performance on easy items than on harder ones. This implies a predicted scoring pattern for each patient on all items (in our scale, a four-level choice was given for every item). The closeness of the observed scores to the predicted scoring pattern is expressed by two fit statistics: (1) the outlier-sensitive fit statistic (Outfit), which is sensitive to unexpected behavior affecting responses to items far from a patient's ability level; and (2) the information-weighted fit statistic (Infit), which is sensitive to unexpected behavior affecting responses to items matching patient ability, and thus most informative of the patient himself or herself. Moreover, conventional point-biserial correlation coefficients (PtBis) are computed for each item, indicating the extent to which patient scores on an item correlate with patient scores on the whole test.

Reliability. Reliability is conventionally defined as the consistency of scores obtained by the same patient (or item) on different occasions, or by different evaluators on the same occasion, or under different examining conditions. It is determined by variance across repeated testing, and computed as the ratio between true measure variance and observed (true + error) measure variance.[25] In Rasch theory, error measure variance is directly computed from the measurement error accompanying each patient ability and item difficulty estimates.[18,26] Repeated testing is not required to compute reliability. A useful related index is separation, defined as the ratio between true spread of the measures (as expressed by their standard deviation corrected for measurement error) and measurement error.[18,26] Separation can be used to estimate the number of strata that are significantly distinguished within the range of observed item difficulties (item separation) and patient abilities (patient separation).

RESULTS

ABILHAND Calibration

The analysis of the 57 ABILHAND items showed that some of them fit the model poorly, as reported by Infit and/or Outfit mean-squares. Both of these fit statistics are expected to approach 1, with typical acceptable values between 0.6 and 1.4.[27] High mean-squares indicate noisy response patterns, whereas lower values represent a muted, overly orderly response pattern. Of the original 57 items, 11 had either Infit or Outfit mean-squares higher than 1.4. This suggests that these items do not pertain to the same underlying domain as the other 46, and are therefore not useful for this measure of manual ability. These items were deleted from subsequent analyses. Items with low mean-squares are potentially redundant, but were maintained in this preliminary analysis for the purpose of further refinement.

The remaining 46 properly fitting items are presented in table 1 in order of decreasing difficulty (range: 3.8 to −3.7 logits). Standard errors (SE) on these difficulties range from 0.4 to 0.8.

Table 1: ABILHAND Items Calibrations and Fit of Data to the Rasch Model

	Items	Usually Bimanual	Difficulty (logits)	SE (logits)	Outfit (Mnsq)	Infit (Mnsq)	PtBis
1.	Screwing a nut on		3.8	0.5	0.8	0.8	0.7
2.	Threading a needle	x	3.2	0.5	1.1	1.1	0.4
3.	Using a screwdriver		2.9	0.5	1.1	1.0	0.6
4.	Hammering a nail	x	2.7	0.5	1.4	1.4	0.8
5.	Handling a stapler		2.5	0.7	1.4	1.3	0.5
6.	Shelling hazel nuts	x	2.5	0.8	0.3	0.3	0.9
7.	Filing one's nails		1.9	0.5	1.2	1.2	0.7
8.	Tearing open a pack of chips	x	1.8	0.5	1.1	1.1	0.6
9.	Buttoning up trousers	x	1.8	0.5	1.0	1.0	0.8
10.	Fastening a snap (jacket, bag, etc.)	x	1.6	0.4	1.7	1.7	0.4
11.	Winding up a wristwatch		1.6	0.5	1.5	1.4	0.7
12.	Drawing		1.5	0.7	1.1	1.1	0.7
13.	Cutting meat	x	1.4	0.4	1.2	1.2	0.7
14.	Buttoning up a shirt	x	1.4	0.4	1.5	1.5	0.4
15.	Grasping a coin on a table		1.2	0.4	1.2	1.2	0.7
16.	Turning a key in a keyhole		1.2	0.5	1.1	1.0	0.7
17.	Taking a coin out of a pocket		0.8	0.4	1.1	1.1	0.7
18.	Making pancake batter	x	0.7	0.8	0.6	0.6	0.3
19.	Typewriting	x	0.7	0.8	0.3	0.3	1.0
20.	Fastening the zipper of a jacket	x	0.7	0.4	1.6	1.6	0.4
21.	Peeling onions	x	0.4	0.5	1.2	1.1	0.7
22.	Counting bank notes	x	0.2	0.6	1.2	1.2	0.7
23.	Handling a 4-color ballpoint pen with one hand		0.0	0.7	1.0	1.0	0.8
24.	Using a spoon		−0.1	0.4	1.0	0.9	0.7
25.	Pulling up the zipper of trousers	x	−0.1	0.5	1.1	1.1	0.9
26.	Combing one's hair		−0.3	0.6	1.2	1.2	0.8
27.	Squeezing toothpaste on a toothbrush	x	−0.5	0.7	0.8	0.7	0.8
28.	Inserting a diskette into a drive		−0.6	0.6	1.2	1.0	0.7
29.	Picking up a can		−0.7	0.4	0.6	1.1	0.7
30.	Closing a door		−0.9	0.5	1.2	1.2	0.5
31.	Spreading butter on a slice of bread	x	−1.0	0.6	0.5	0.5	0.8
32.	Brushing one's hair		−1.1	0.7	0.7	0.6	0.7
33.	Brushing one's teeth		−1.1	0.6	1.3	1.1	0.7
34.	Wrapping up gifts	x	−1.4	0.4	1.5	1.3	0.6
35.	Washing one's hands	x	−1.4	0.4	0.9	0.8	0.7
36.	Eating a sandwich		−1.8	0.6	0.7	0.6	0.7
37.	Dialing on a keypad phone		−1.9	0.8	0.1	0.1	1.0
38.	Placing a glass of water on a table		−2.2	0.6	0.1	0.1	0.9
39.	Unwrapping a chocolate bar	x	−2.2	0.5	0.7	0.7	0.6
40.	Drinking a glass of water		−2.3	0.5	0.4	0.4	0.8
41.	Turning on a radio		−2.5	0.5	0.6	0.5	0.6
42.	Ringing a door bell		−2.5	0.5	0.9	0.9	0.6
43.	Washing one's face		−2.5	0.6	0.2	0.2	0.8
44.	Blowing one's nose		−2.8	0.6	0.3	0.2	0.8
45.	Turning on a lamp		−2.9	0.5	0.9	0.7	0.6
46.	Turning on a television set		−3.7	0.5	1.0	1.0	0.3
Mean			0.0	0.5	0.9	0.9	0.7
SD			1.9	0.1	0.4	0.4	0.2

This 46-item set fits the model unidimensionality specification, and their fit mean-squares globally approximate the expected unitary value. This indicates that within this preliminary calibration, these items contribute to the definition of a single underlying ability construct. Further support is provided by all PtBis being positive (minimum 0.3).

Patient abilities and item difficulties are mapped in figure 1. In Rasch measurement, patient measures and item difficulties are calibrated along the same scale—the measure of the underlying ability, which is represented vertically on the graph. The figure shows the useful range covered by ABILHAND when considering the ability required to answer in all four response categories (very difficult to very easy) to all items. In the column labeled Mean, each item is plotted at its mean difficulty as listed in table 1. In the column labeled Step 1, each item is plotted at the ability level corresponding to an expected score of 0.5 on the response scale. In the column labeled Step 3, each item is plotted at the ability level corresponding to an expected score of 2.5 on the response scale. Thus, the useful range of the questionnaire completely overlaps our 18 patient measures, indicating that item difficulties cover the patient abilities in the sample tested.

The relationship between ordinal raw scores on the questionnaire and the corresponding interval ABILHAND measure is

```
ABILHAND (logits)

   |---------------|----------------------------------------------|
   |   Patient     |              Item   difficulties             |
   |   abilities   |   Step  1   |      Mean       |   Step  3  |
 8 +---------------+-------------+-----------------+------------+  8
   | More able     |             |                 |     Harder |
 7 +               +             +                 + X          +  7
   |               |             |                 |            |
 6 +               +             +                 + XX         +  6
   |               |             |                 | XXX        |
 5 +             O +             +                 + XXX        +  5
   |               |             |                 | XXXXX      |
 4 +             O +             + X               + XXX        +  4
   |             O |             |                 | XXXX       |
 3 +               +             + XX              + XXXX       +  3
   |               |             | XXX             | XXXX       |
 2 +             O +             + XXX             + XXXX       +  2
   |            OO |             | XXXXX           | XX         |
 1 +               +             + XXX             + XXXXX      +  1
   |           OOO | X           | XXXX            | XXX        |
 0 +             O +             + XXXX            + XX         +  0
   |         OOOOO | XX          | XXXX            | X          |
-1 +           OOO + XXX         + XXXX            +            + -1
   |               | XXX         | XX              |            |
-2 +               + XXXXX       + XXXXX           +            + -2
   |               | XXX         | XXX             |            |
-3 +               + XXXX        + XX              +            + -3
   |               | XXXX        | X               |            |
-4 +               + XXXX        +                 +            + -4
   |               | XXXX        |                 |            |
-5 +               + XX          +                 +            + -5
   |               | XXXXX       |                 |            |
-6 +               + XXX         +                 +            + -6
   |               | XX          |                 |            |
-7 +               + X           +                 +            + -7
   | Less able     |             |                 |     Easier |
-8 +---------------+-------------+-----------------+------------+ -8
```

Fig 1. Map of patient abilities and item difficulties along the ABILHAND scale. In the left column each "o" represents one patient (bold for bilaterally operated) along the ability axis going from less able (bottom) to more able (top). On the right side, each item is represented three times by an "x." In the column labeled Mean, each item is plotted at its mean difficulty (bold symbols for bimanual activities) on the axis of the underlying variable going from easier (bottom) to harder items (top). In the column labeled Step 1, each item is plotted at the ability level corresponding to the transition from a score of 0 (very difficult) to 1 (slightly difficult) on the rating scale, indicating the lowest difficulty level of each item. In the column labeled Step 3, each item is plotted at the ability level corresponding to the transition from a score of 2 (easy) to 3 (very easy), indicating the highest difficulty of each item.

plotted in figure 2. The ogival relationship indicates that the raw scale is approximately linear between 45 and 95, within the 0 to 138 potential range. Outside this central range, however, each score point corresponds to an increasing amount of ability. The 95% confidence interval shows the minimum statistical difference between scores required to represent different levels of ability. Within this preliminary calibration, this minimum difference is about 16 score points in the central part of the scale, ranging from 0 to 138.

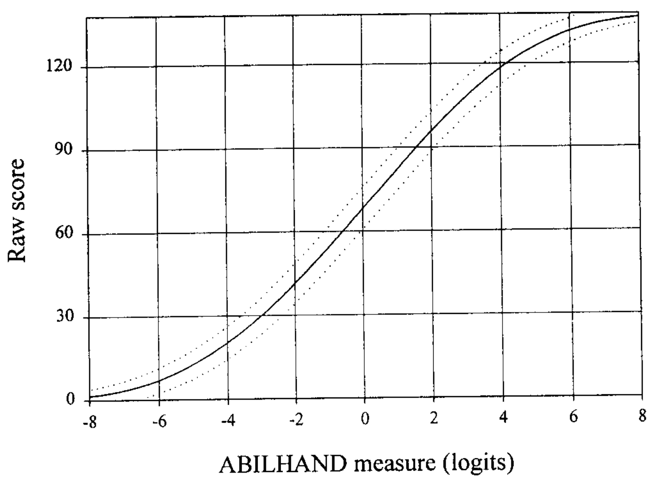

Fig 2. Ogival relationship between finite raw scores (range: 0 to 138) on the ordinate and the infinite ABILHAND measure (−8/+8 logits range plotted) on the abscissa. The dotted lines indicate the 95% confidence interval.

Reliability

The hierarchical structure of the test, expressed by item separation, indicates that items are distributed into 4.3 difficulty strata, leading to an item separation reliability of .90. On the other hand, patient reliability equals .96, indicating that 6.9 statistically different levels of ability can be distinguished in this sample.[26]

DISCUSSION

The ABILHAND items were selected to explore a range of manual activities and to define a common underlying trait of manual ability. Items could be perceived as difficult if they required remarkable sensorimotor function, like grip strength or dexterity (eg, items 1, 9, and 11), implied a heavy mental effort (eg, items 12 and 19), were tiring (eg, items 13 and 22), or involved risk of breakage (eg, item 38). In spite of these potential idiosyncrasies in item content and perceived difficulty, all 46 items fit the model within this preliminary calibration, and therefore contribute their own information to the measurement of a common ability continuum in the sample tested. Patients made up a highly selected, homogeneous sample, presenting quite a rare condition. Despite the low number of patients, the Rasch method allowed an effective refinement of the scale by driving the process of collapsing categories and verifying scale unidimensionality. This version of ABILHAND seems to be a good starting point for further refinements based on more powerful analyses of more frequent impairments (eg, stroke patients).

The Rasch-derived scale is consistent with clinical experience. For instance, Tennant and associates[28] report that for the Stanford Health Assessment Questionnaire (HAQ),[29,30] the item "lifting a full cup or glass to the mouth" is the easiest one for patients with RA. Similarly in our questionnaire, "drinking a glass of water" and "placing a glass of water on a table" are among the easiest items. Six items in our questionnaire are still easier, but none of them are included in the HAQ. Furthermore, Heinemann and colleagues[31] showed that the Functional Independence Measure item "feeding" requires less ability than "grooming," which in turn is easier than "dressing upper body" for patients with RA. Similarly, when regrouping items relating to feeding (ie, items 13, 24, 29, 31, 36, 38, and 40), grooming (ie, items 7, 26, 27, 32, 33, 35, and 43), and dressing the upper body (ie, items 10, 14, and 20) in the ABILHAND, we observe that the mean difficulties equal −.94, −.71, and 1.23 logits, respectively, indicating that the same hierarchy is reproduced. This also reinforces the comparability between an examiner's standardized rating as implemented in the FIM and the self-estimation by a patient as in the ABILHAND.[19,32]

In addition, ABILHAND includes items that were found to be quite difficult for our sample, and related to classic "extended" activities of daily living, such as cooking, office work, and handiwork. This extension of the measurement range beyond the basic activities of daily living is desirable in following patients in their lives or job environments.

ABILHAND also confirms that the Rasch method can provide clinicians with powerful help in developing and validating measurement scales.

Acknowledgments: We are grateful to John-Michael Linacre, University of Chicago, Alan Tennant, University of Leeds, and Norman Heglund, Catholic University of Louvain, for reviewing this paper. We also extend our appreciation to Jean-Jacques Rombouts, Catholic University of Louvain, who allowed us to evaluate his patients.

References

1. Jones LA. The assessment of hand function: a critical review of techniques. J Hand Surg [Am] 1989;14:221-8.

2. Thonnard JL, Plaghki L, Bragard D. Evamain: computerized system for the evaluation of the hand. In: Scuind F, An KN, Cooney WP, Garcia-Elias M, editors. Advances in the biomechanics of the hand and wrist. New York: Plenum Press; 1994. p. 499-510.
3. Carroll D. A quantitative test of upper extremity function. J Chron Dis 1965;18:479-91.
4. De Souza LH, Hewer RL, Miller S. Assessment of recovery of arm control in hemiplegic stroke patients. 1. Arm function tests. Int Rehabil Med 1980;2:3-9.
5. Desrosiers J, Hébert R, Bravo G, Dutil E. Upper extremity performance test for the elderly (TEMPA): normative data and correlates with sensorimotor parameters. Arch Phys Med Rehabil 1995;76:1125-9.
6. Heller A, Wade DT, Wook VA, Sunderland A, Hewer RL, Ward E. Arm function after stroke: measurement and recovery over the first three months. J Neurol Neurosurg Psychiatry 1987;50:714-9.
7. Jebsen RH, Taylor N, Trieschmann RB, Trotter MJ, Howard LA. An objective and standardized test of hand function. Arch Phys Med Rehabil 1969;50:311-9.
8. Lindmark B, Hamrin E, Törnquist K. Testing daily functions post-stroke with standardized practical equipment. Scand J Rehabil Med 1990;22:9-14.
9. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. Int J Rehabil Res 1981;4:483-92.
10. Potvin AR, Tortelotte WW, Dailey JS, Alberts JW, Walker JE, Pew RW, et al. Simulated activities of daily living examination. Arch Phys Med Rehabil 1972;53:476-86.
11. Wilson DJ, Baker LL, Craddock JA. Functional test for the hemiparetic upper extremity. Am J Occup Ther 1984;38:159-64.
12. World Health Organization. International classification of impairments, disabilities, and handicaps. Geneva: World Health Organization; 1980.
13. Bach-y-Rita P. Brain plasticity as a basis of the development of rehabilitation procedures for hemiplegia. Scand J Rehabil Med 1981;13:73-83.
14. Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. Arch Phys Med Rehabil 1989;70:308-12.
15. Wright BD, Linacre JM. Observations are always ordinal; measurement, however, must be interval. Arch Phys Med Rehabil 1989;70: 857-60.
16. Fisher AG. Functional measures, part 2: selecting the right test, minimizing the limitations. Am J Occup Ther 1992;46:278-81.
17. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: Mesa Press; 1992.
18. Wright BD, Masters GN. Rating scale analysis. Chicago: Mesa Press; 1982.
19. Myers AM, Holliday PJ, Harvey KA, Hutchison KS. Functional performance measures: are they superior to self-assessments? J Gerontol 1993;48:M196-206.
20. Tesio L. Disability, dependence and performance: which is which? Europa Medicophysica 1997;33:55-7.
21. Mannerfelt L, Malmsten M. Arthrodesis of the wrist in rheumatoid arthritis. A technique without external fixation. Scand J Plast Reconstr Surg 1971;5:124-30.
22. Linacre JM. Categorical misfit statistics. Rasch Measurement Trans 1995;9:450-1.
23. Andrich D. Category ordering and their utility. Rasch Measurement Trans 1996;9:464-5.
24. Linacre JM, Wright BD. BIGSTEPS: Rasch model computer program. Chicago: Mesa Press; 1994.
25. Anastasi A. Psychological testing. 6th ed. New York: Maxwell Macmillan International; 1990.
26. Fisher WP. Reliability statistics. Rasch Measurement Trans 1992;6: 238.
27. Wright BD, Linacre JM. Reasonable mean-square fit values. Rasch Measurement Trans 1994;8:370.
28. 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.
29. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137-45.
30. Kirwan JR, Reeback JS. Stanford Health Assessment Questionnaire modified to assess disability in British patients with rheumatoid arthritis. Br J Rheumatol 1986;25:206-9.
31. Heinemann AW, Hamilton BB, Granger CV, Wright BD, Linacre JM, Betts HB, et al. Rating scale analysis of functional assessment measures. Final report to the National Institute on Disability and Rehabilitation Research. Chicago: Rehabilitation Institute of Chicago; 1991.
32. Brown RG, MacCarthy B, Jahanshahi M, Mardsen CD. Accuracy of self-reported disability in patients with parkinsonism. Arch Neurol 1989;46:955-9.