

ORIGINAL REPORT

REPRODUCIBILITY OF THREE SELF-REPORT PARTICIPATION MEASURES: THE ICF MEASURE OF PARTICIPATION AND ACTIVITIES SCREENER, THE PARTICIPATION SCALE, AND THE UTRECHT SCALE FOR EVALUATION OF REHABILITATION-PARTICIPATION

Carlijn H. van der Zee, MSc¹, Annique R. Priesterbach, BSc¹, Luikje van der Dussen, MD², Albert Kap, MD¹, Vera P. M. Schepers, MD PhD¹, Johanna M. A. Visser-Meily, MD PhD^{1,3} and Marcel W. M. Post, PhD^{1,3}

From the ¹Centre of Excellence for Rehabilitation Medicine, Rehabilitation Centre De Hoogstraat, ²Rehabilitation Centre De Trappenberg, Location Almere and ³Rudolf Magnus Institute of Neuroscience, Department of Rehabilitation, Nursing Science and Sports, University Medical Centre Utrecht, The Netherlands

Objective: To assess the reproducibility of 3 participation measures.

Design: Repeated administration of a postal questionnaire with a 2-week interval.

Participants: Outpatients ($n=47$) from 2 rehabilitation centres and a university hospital in The Netherlands.

Methods: Measures were the ICF Measure of Participation and Activities Screener (IMPACT-S), the Participation Scale, and the Utrecht Scale for Evaluation of Rehabilitation-Participation (USER-P). Test-retest reliability was analysed using Cohen's weighted kappa and the intraclass correlation coefficient (ICC). Agreement was expressed as the standard error of measurement and the smallest detectable change (SDC), substantiated as the ratio between the SDC and the standard deviation (SDC/SD).

Results: ICC values of the IMPACT-S were 0.54–0.90 for the scale scores, 0.92 and 0.74 for sub-total scores Activities and Participation, and 0.88 for the total score. The ICC of the Participation Scale was 0.82. The ICC of the USER-P was 0.65 for the Frequency scale, 0.85 for the Restrictions scale, and 0.84 for the Satisfaction scale. The SDC/SD ratios for all measures were small (0.11–0.28) at the group level, but large (0.78–1.91) at the individual level. Most participants found all measures relevant and easy to complete.

Conclusion: All 3 measures showed generally satisfying reproducibility and were acceptable to the participants.

Key words: reproducibility of results; validation studies; community participation; outcome measure.

J Rehabil Med 2010; 42: 752–757

Correspondence address: Marcel Post, Rembrandtkade 10, 3583 TM Utrecht, The Netherlands. E-mail: m.post@dehoogstraat.nl

Submitted October 15, 2009; accepted May 26, 2010

INTRODUCTION

Most patients are referred to rehabilitation because of conditions that cannot be cured. Their treatment will be aimed at minimizing the consequences of these conditions to improve

independence and, ultimately, social participation (1). In the outpatient clinic in particular, re-establishment of social participation is a key aim of rehabilitation programmes. Measurement of participation outcomes is, however, not common in rehabilitation research (2, 3). This discrepancy has been related to the nature of participation as being affected by many factors outside the control of the rehabilitation team, but also to measures of participation being less developed than measures of more basic activities (4). Since the introduction of the International Classification of Functioning, Disability and Health (ICF) in 2001 (5), many instruments to measure participation have been developed, but psychometric evidence on these measures is still incomplete (3). A participation measure, like any measure, must be valid, reproducible, and responsive in order to be used as an outcome measure (6). Existing participation measures have generally showed validity, but their reproducibility and responsiveness have rarely been established (2, 3, 7).

In response to this lack of data, we started a prospective multi-centre study to identify a valid and responsive instrument to measure participation outcomes of outpatient rehabilitation (8). Participation measures were selected for this study using the following criteria: (i) applicable in various diagnostic groups; (ii) feasible (being brief and suitable for self-report) for use in routine outcome monitoring; (iii) providing both objective and subjective ratings of participation; (iv) covering the ICF participation chapters (5); and (v) having sound psychometric properties. No measure met all criteria, but we identified several promising measures, 4 of which were selected for our responsiveness study (8). The Frenchay Activities Index (FAI) (9) was selected because it is the most often used participation measure in rehabilitation research (3), and the only participation measure used in clinical practice in The Netherlands. The ICF Measure of Participation and Activities Screener (IMPACT-S) (10) was selected because it is the only participation measure that covers all Activities and Participation chapters of the ICF (5). It is a measure we developed in earlier research (8). The Participation Scale (11) was selected

because it is the only participation measure that asks people to rate their participation using an explicit frame of reference, namely the “peer group”. Finally, since we found no instrument measuring both objective and subjective participation and which satisfied most other criteria, we developed a new measure, the Utrecht Scale for Evaluation of Rehabilitation-Participation (USER-P) (8). Our study into the responsiveness of these 4 measures is ongoing. However, except for the FAI (12), there was also a need for data on the reproducibility of these measures. The reproducibility of the IMPACT-S has been studied previously (10), but some alterations have been made to this measure since then. Evidence of the reproducibility of the Participation Scale is incomplete and, to include this scale in our responsiveness study, we had to translate it into Dutch and transform it from an interviewer-administration into a self-report measure so that the reproducibility of this Dutch self-report version also had to be assessed. The same was true for the USER-P as a newly developed measure. The aim of the present study was therefore to assess the reproducibility of the IMPACT-S, the USER-P, and the Participation Scale.

METHODS

Sample

A total of 104 candidate-participants with physical disabilities were selected from the outpatient clinics of rehabilitation centres De Hoogstraat and De Trappenberg, in Almere, and the University Medical Centre Utrecht, The Netherlands. Inclusion criteria were a minimum age of 18 years and the ability to read and comprehend self-report measures in Dutch. Exclusion criteria were severe cognitive impairments, aphasia, and a rapidly progressive disorder.

Procedure

Candidate-participants received a written invitation to participate in the study along with the questionnaire. Participants who did not respond within two weeks received a once-only reminder. Participants who replied with a completed questionnaire, received the second questionnaire two weeks after completing the first. Participants who did not return this second questionnaire within two weeks received a reminder. The study protocol was approved by the local medical ethics board of Rehabilitation Centre De Hoogstraat.

Instruments

The IMPACT-S, the Participation Scale, and the USER-P were combined in random order in the questionnaire and it was ensured that participants would receive the measures in different order on both administrations. In addition to these measures, the first questionnaire contained questions on diagnosis and demographic characteristics and the second questionnaire contained questions on the respondent's opinion about the measures, asking for the most relevant and easiest measure as well as asking for possible irrelevant or obtrusive questions.

The IMPACT-S assesses experienced limitations in activities and participation comprising 32 items covering all 9 chapters of the Activities and Participation component of the ICF (5). All items are rated on a score of 0 (cannot do that at all) to 3 (no limitations whatsoever). Nine scale scores, two sub-total scores for Activities and Participation and a total score can be computed. All summary scores are converted to a score on a 0–100 scale, in which a high score indicates a high level of participation. The test-retest reliability of IMPACT-S has been assessed in road accident victims and was found to be good at item level ($\kappa=0.44-0.72$), scale level (Intraclass Correlation Coefficient

(ICC)=0.72–0.92), sub-total score level (0.90–0.93), and total score level (0.94) (10). However, after finishing this study one item was omitted from this measure and the number of response options has been increased from 3 to 4 because it was expected that separating the previously merged categories “considerable limitations” and “I cannot do that at all” would make it easier for respondents to choose the category that best reflects their situation.

The Participation Scale measures experienced participation restrictions (11). It covers 8 out of 9 ICF Activities and Participation chapters. Originally the Participation Scale was an interview-based instrument. It was translated into Dutch and re-designed as a self-report measure in co-operation with the author. The Participation Scale contains 18 items, each measuring the level of participation compared with peers and, in case of a lower level of participation, the extent to which the respondent experiences this as a problem. “Peers” are defined as: people who are similar to the respondent in all aspects (socio-cultural, economic, and demographic) except for the health condition or disability (13). Both answers are combined in an item score between 0 (same level of participation) and 5 (lower level of participation and experienced as a large problem). A total Participation Scale score is obtained as the sum of the item scores, ranging from 0 to 90, with a high score indicating severe participation restrictions. The Participation Scale was found to be valid and reliable, with a Cronbach's α of 0.92, a test-retest reliability ICC of 0.83, and inter-tester reliability of 0.80 (11).

The USER-P is a newly developed participation measure that aims to measure both objective and subjective participation. It is an extension of the USER, which is a measure of activity limitations (14). The USER-P consists of 31 items, covering 8 out of 9 ICF Activities and Participation chapters. It assesses 3 aspects of participation: frequency, experienced restrictions, and satisfaction. (i) Frequency of participation consists of two parts: the first part contains 4 items on frequency of vocational activity measuring the amount of time the respondent spends on paid work, unpaid work, study, and housekeeping in a typical week. Each item is scored from 0 (not at all) up to 5 (36 hours or more). The second part contains 8 items on frequency of leisure and social activity measuring the frequency of performing activities in the past 4 weeks such as visiting family or friends. Each item is scored from 0 (not at all) to 5 (19 times or more), with higher scores reflecting higher levels of participation. (ii) Participation restrictions are assessed by asking the respondent for experienced restrictions as a result of his/her health condition in 10 activities, such as making day-trips and other outdoor activities. Each item score ranges from 0 (not possible at all) to 3 (no difficulty at all), with a higher score indicating less participation restrictions. (iii) Satisfaction with participation is determined by asking the respondent to indicate the satisfaction with 9 aspects of life, such as contacts with family members. Items are rated on a scale of 0 (not satisfied at all) to 4 (very satisfied), with a higher score indicating more satisfaction. The sum scores for the Frequency, Restrictions, and Satisfaction scales are all converted to scores on a 0–100 scale. There is no USER-P total score.

Statistical analyses

Data were analysed using SPSS 16.0. Floor and ceiling effects were considered present if 15% of respondents scored, respectively, the lowest or highest score on a scale (15). The skewness of the score distribution was assessed and considered acceptable if the skewness was between -1 and 1 . Parametric tests to assess reproducibility were used since almost all scores were normally distributed and there are no non-parametric alternatives for these tests.

Reproducibility is the extent to which similar scores are obtained on repeated administration of a measure when no substantial change has occurred in the time between the measurements. Reproducibility consists of two different, but related, aspects: reliability and agreement (6). Reliability concerns the degree to which patients can be distinguished from each other despite measurement error. The test-retest reliability on item level was analysed using Cohen's weighted kappa. A weighted kappa of 0.21–0.40 was considered fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1.00 almost perfect (16). The test-retest reliability

on the level of scale scores, sub-total scores, and total scores were examined using the ICCs, using the model for absolute agreement (6). An ICC was considered satisfactory if above 0.75 (17).

Agreement concerns the absolute measurement error, i.e. how similar scores on repeated administrations are, expressed in the unit of the measurement scale at issue. Small measurement error is required for evaluation purposes, in which one wants to distinguish clinically important change from measurement error (6). Agreement was analysed using the standard error of measurement (SEM) and the smallest detectable change (SDC). The SEM equals the square root of the error variance, including systematic differences (6). The SEM was considered small if it represented less than 10% of the score range (18). The SEM can be converted into the smallest detectable change (SDC_{ind}) by multiplying the SEM by $1.96\sqrt{2}$. The SDC_{ind} reflects the smallest change in score of an individual that can be interpreted as "real" change, i.e. change above measurement error at an alpha level of 0.05 (6). To determine the SDC on group level (SDC_{group}), the SDC_{ind} is divided by \sqrt{n} (6). To assess responsiveness, the SDC should ideally be compared with the score difference representing clinically relevant change. However, this figure is not available for the measures under study. Alternatively, we used the ratio of the SDC and the average standard deviation (SD) of the scores on both measurements to substantiate the SDC. An SDC/SD of more than 0.8 was interpreted as requiring large score differences to exceed chance (19).

RESULTS

A total of 104 individuals (42 men, 62 women) were invited to participate in this study, of which 47 individuals participated in both measurements. Three individuals only completed the first questionnaire and 54 individuals did not participate at all. The response of the males (33%) was somewhat lower than that of the females (52%). The mean age of the non-responders and responders was similar, with 50.9 (SD=14.5) and 50.6 (SD=11.8) years, respectively. Response rate was significantly related to diagnosis, from 34% in patients with a stroke up to 72% in patients with a musculoskeletal condition. Participants' characteristics are shown in Table I.

Psychometric properties of each measure are displayed in Table II. The main findings are summarized for each measure separately.

IMPACT-S

The proportion of missing item responses was small (1.1%). The means and medians of all IMPACT-S scores were high, considering the score range. All scale scores of the IMPACT-S, except for the scale Mobility, showed ceiling effects (range 15.2–55.3%). The sub-total scores for Activities and Participation and the total score did not show floor or ceiling effects (data not displayed). The skewness was acceptable for all scores, except the scale score Community (–1.03), the sub-total score Participation (–1.45), and the total score (–1.09). The mean percentage of exact agreement between individual items on the two measurements was 73.1% (range 56.5–89.1%). Weighted kappa values for the individual items ranged from 0.22 to 0.82 and were fair for 3 items, moderate for 7 items, substantial for 19 items, and almost perfect for 3 items.

Differences between scores on the first and second measurements were small (Table II). Three out of 9 scale scores, the sub-total scores, and the total score showed satisfactory ICC

Table I. Participants' characteristics

Characteristics	
Gender, <i>n</i> (%)	
Men	15 (32)
Women	32 (68)
Mean age, years (SD)	50.6 (11.8)
Diagnosis, <i>n</i> (%)	
Musculoskeletal disease	8 (17.0)
Traumatic brain injury	5 (10.6)
Stroke	12 (25.5)
Neuromuscular diseases	11 (23.4)
Chronic pain	10 (21.3)
Heart failure	1 (2.1)
Median time since diagnosis, years (range)	1.7 (0.7–15.6)
Paid job before condition, <i>n</i> (%)	
Yes	31 (66)
No, reason:	
Housekeeping	4 (8.5)
Retirement	4 (8.5)
Student	2 (4.3)
Health problems	4 (8.5)
Other	2 (4.3)
Current marital status, <i>n</i> (%)	
Married/living together	34 (72.3)
Other	13 (27.7)
Education, <i>n</i> (%)	
Lower	19 (40.4)
Higher	27 (57.4)
Median time between measurements, days (range)	16.0 (13–49)

SD: standard deviation.

values (Table II). The SEM was below 10% of the score range for all scores, except for 3 out of 9 scale scores. The SDC_{ind}/SD ratio was above 0.8 for all scores, except for the sub-total score Activities of the IMPACT-S. The SDC_{group}/SD ratio was small for all scores (range 0.11–0.28).

Participation scale

The proportion of missing item responses was somewhat larger than that for the other measures (2.8%). The mean and median scores of the Participation Scale were low (indicating less participation restrictions), considering the score range. The skewness of this score was acceptable and there were no floor or ceiling effects. The mean percentage of exact agreement between individual items was 70.3% (range 51.5–93.2%). Weighted kappa values of the individual items ranged from 0.00 to 0.87 and were slight for 2 items, moderate for 5 items, substantial for 8 items, and almost perfect for 3 items. The Participation Scale showed a satisfying ICC (Table II). Agreement expressed by the SEM was well below 10% of the score range. The SDC_{ind}/SD ratio was above 0.8 and the SDC_{group}/SD was small.

USER-P

The proportion of missing item responses was small (1.3%). The mean and median scores on the Restriction scale were fairly high, considering the score range. The skewness of all scales was acceptable and there were no floor or ceiling ef-

Table II. Reproducibility of the IMPACT-S, Participation Scale, and USER-P (n=47)

	Score range	Median (IQR)	Mean (SD)	Difference test-retest M (SD)	ICC (95% CI)	SEM	SDC _{ind}	SDC _{ind} /SD	SDC _{group}	SDC _{group} /SD
IMPACT-S Total	0-100	82.3 (72.9-88.5)	79.6 (12.3)	0.3 (6.2)	0.88 (0.80-0.93)	4.4	12.1	0.96	1.8	0.14
Activities	0-100	80.4 (70.4-88.9)	78.4 (12.6)	0.4 (4.9)	0.92 (0.86-0.93)	3.5	9.7	0.78	1.4	0.11
Participation	0-100	83.3 (75.6-88.1)	81.1 (14.5)	0.4 (10.9)	0.74 (0.58-0.92)	7.7	21.3	1.42	3.1	0.21
Knowledge	0-100	77.8 (66.7-88.9)	79.7 (15.6)	1.4 (12.6)	0.70 (0.52-0.82)	8.9	24.7	1.52	3.6	0.22
General tasks	0-100	83.3 (50.0-100.0)	72.3 (23.9)	2.8 (19.8)	0.67 (0.48-0.80)	14.0	38.7	1.59	5.6	0.23
Communication	0-100	88.9 (66.7-100.0)	84.8 (17.5)	-0.4 (9.7)	0.85 (0.74-0.91)	6.9	19.1	1.09	2.8	0.16
Mobility	0-100	71.4 (61.9-85.7)	71.8 (19.7)	0.0 (8.9)	0.90 (0.84-0.95)	6.3	17.4	0.86	2.5	0.12
Self-care	0-100	88.9 (77.8-100.0)	90.2 (11.5)	-0.8 (11.0)	0.58 (0.36-0.74)	7.8	21.5	1.80	3.1	0.26
Domestic life	0-100	83.3 (66.7-91.7)	76.3 (18.9)	-0.4 (12.2)	0.80 (0.68-0.89)	8.6	23.8	1.23	3.5	0.18
Relationships	0-100	91.7 (83.3-100.0)	88.0 (15.8)	-0.4 (11.9)	0.68 (0.49-0.81)	8.4	23.3	1.58	3.4	0.23
Major life areas	0-100	83.3 (66.7-100.0)	78.6 (19.5)	4.0 (18.3)	0.54 (0.30-0.71)	13.1	36.4	1.91	5.3	0.28
Community	0-100	83.3 (66.7-93.8)	80.3 (18.5)	0.2 (17.9)	0.59 (0.37-0.75)	12.6	35.0	1.79	5.1	0.26
Participation Scale	0-90	15.0 (6.0-27.0)	17.1 (12.9)	-0.3 (7.9)	0.82 (0.70-0.90)	5.6	15.5	1.18	2.3	0.17
USER-P										
Frequency	0-100	32.5 (27.5-38.8)	33.8 (9.8)	0.7 (7.8)	0.65 (0.45-0.79)	5.5	15.2	1.65	2.2	0.24
Restrictions	0-100	73.3 (63.0-80.0)	70.6 (17.9)	-1.8 (9.2)	0.85 (0.75-0.92)	6.6	18.2	1.07	2.7	0.16
Satisfaction	0-100	63.9 (50.0-77.8)	63.6 (17.3)	-1.5 (9.7)	0.84 (0.73-0.91)	6.8	18.9	1.12	2.8	0.16

Mean and SD are shown for first measurement. Note that high scores on the IMPACT-S and USER-P indicate high levels of participation and that a high score on the Participation Scale indicates large participation restrictions.

IQR: interquartile range; SD: standard deviation; M: mean; CI: confidence interval; SEM: standard error of measurement; SDC: smallest detectable change; IMPACT-S: ICF Measure of Participation and Activities Screener; USER-P: Utrecht Scale for Evaluation of Rehabilitation-Participation.

fects. Exact agreement between the items was 67.2% (range 39.1-95.3%). Weighted kappa values of the individual items ranged from 0.30 to 0.95 and were fair for 2 items, moderate for 9 items, substantial for 13 items, and almost perfect for 7 items. The differences between mean scores on the first and second measurements were very small. The Restrictions and the Satisfaction scales showed satisfying reliability, but the Frequency scale showed less than satisfying ICC values. Agreement expressed by the SEM was well below 10% of the score range. The SDC_{ind}/SD ratio was above 0.8 and the SDC_{group}/SD was small.

Respondent's opinion

More than half of the respondents considered all 3 measures to be a relevant measure for their participation. Respondents who preferred one of the measures judged the USER-P to be the best and the Participation Scale to be the least favourable. One respondent found none of the measures relevant for measuring participation. Furthermore, more than half of the respondents considered all 3 measures to be easy to complete. Most respondents who preferred one of the measures, found the USER-P the easiest and the Participation Scale the least easy to complete. Four respondents found none of the measures easy. A common comment concerned the layout of the Participation Scale, which was perceived as confusing. Few respondents mentioned obtrusive items, but it was mentioned that items confronted them with their restrictions and that items regarding partner relationship were frustrating, especially when they were single due to reasons other than their condition (e.g. widowhood).

DISCUSSION

This study showed generally satisfactory reproducibility of all 3 measures. The SDC was small at the group level, but large at the individual level for all measures, which means that at individual level, large score differences are required to exceed change, while at group level, small score differences will already exceed change. This study adds to the literature by providing psychometric evidence on 3 recently developed participation measures. Agreement figures of the Participation Scale have not been published previously, and this is the first replication of psychometric evidence for both the self-report version of the Participation Scale and the IMPACT-S after their original publication. Furthermore, this study is the first to report psychometric properties of the USER-P.

The test-retest reliability of the IMPACT-S in this study was lower than in the earlier validation study (10) for most scale scores and the sub-total score Participation (0.74 against 0.90). A possible explanation is that the time since diagnosis in the earlier study was longer so that their respondents might have had a more stable level of participation. The agreement figures of the earlier study, calculated with a slightly different method, namely the Smallest Detectable Difference (SDD), however, were similar to the results found in our current study (10).

The test-retest reliability of the Participation Scale in this study was satisfactory and was similar to the figures found in

the earlier validation study (11). Data on agreement are not available, thus no comparison can be made.

This is the first study to assess the reproducibility of the USER-P. Comparisons can therefore only be made with other measures. The reliability of the Frequency scale was lower than that of the Restrictions and the Satisfaction scales. This is consistent with the finding of Brown et al. (20), that the subjective component of the Participation Objective, Participation Subjective instrument showed better reliability coefficients than the objective component. An explanation for this finding might be that actual participation, such as, for example, going to the cinema or doing shopping for fun, is more variable over time than the experience of being restricted in performing these and other activities. The SEM and SDC figures of the Frequency scale were however similar to those of the other USER-P scales. Furthermore, the reliability of the Restrictions scale was similar to the reproducibility of the IMPACT-S (ICC=0.88) (10) and the Dutch version of Life Habits Questionnaire (ICC range 0.78–0.80) (21), which are both also measures of perceived participation restrictions. The test-retest reliability of the Satisfaction score was similar to the ICC of the Personal Wellbeing Index (0.84), which is also a measure of satisfaction with different life domains (22).

Limitations

The sample size in this study was small, but the number of respondents was slightly below the recommended number of 50 (13). Secondly, a heterogenic study sample was used, and the sample size was too small to compute diagnostic group-specific results. However, this heterogeneity of the sample also means that it is easier to generalize the results of this study to the whole population of persons with disabilities treated in outpatient clinics of rehabilitation centres.

Implications for the choice of measure

The 3 participation measures included in this study were selected because they appeared applicable in various diagnostic groups, feasible for use in routine outcome monitoring, covered the ICF participation chapters, and had sound psychometric properties. This study did not reveal a clear “best” among the 3 selected participation measures with respect to reproducibility. However, some differences exist. The separate IMPACT-S scales and the Frequency scale of the USER-P appear less reproducible than the other scores. The IMPACT-S showed a stronger ceiling effect than the other two measures. The USER-P was slightly favoured by the participants. Other differences between these 3 measures are also relevant to make a choice. The Participation Scale, containing 18 items, is the shortest measure, but the length difference will be smaller in practice because each item of the Participation Scale has a follow-up item in case of an experienced restriction. All measures list were suitable for self-report, but the lay-out of the Participation Scale was confusing for some of the respondents. An internet-based version of the questionnaire might solve this problem. All measures cover the participation chapters of the ICF, but the IMPACT-S was the only measure covering all 9 chapters,

while both USER-P and Participation Scale covered 8 out of 9 chapters. Both the IMPACT-S and the Participation Scale measure subjective participation by asking about experienced restrictions. The USER-P measured both objective participation (with the Frequency scale) and subjective participation (with the Restrictions and Satisfaction scales). Potential users of a participation measure can use this information to make a well-informed choice.

REFERENCES

1. Wade DT, de Jong BA. Recent advances in rehabilitation. *BMJ* 2000; 320: 1385–1388.
2. Dijkers MPJM, Whiteneck G, El-Jaroudi R. Measures of social outcomes in disability research. *Arch Phys Med Rehabil* 2000; 81 Suppl 2: S63–S80.
3. Magasi S, Post MW. A comparative review of contemporary participation, measures' psychometric properties and content Coverage. *Arch Phys Med Rehabil* (in press).
4. Wade DT. Outcome measures for clinical rehabilitation trials: impairment, function, quality of life, or value? *Am J Phys Med Rehabil* 2003; 82: S26–S31.
5. World Health Organization. *International Classification of Functioning, Disability and Health*, Geneva: WHO; 2001.
6. De Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006; 59: 1033–1039.
7. Magasi SR, Heinemann AW, Whiteneck GG. Participation following traumatic Spinal Cord Injury: an evidence-based review for research. *J Spinal Cord Med* 2008; 31: 145–156.
8. Van der Zee CH, Post WMW, Kap A, Schouten E, Van Royen R, Rambaran Mishre R, et al. [Measurement of participation as an outcome of outpatient rehabilitation: a prospective multi-centre study]. *Revalidata* 2008; 146: 10–14 (in Dutch).
9. Holbrook M, Silbeck CE. An activities index for use with stroke patients. *Age Aging* 1983; 12: 166–170.
10. Post MWM, De Witte LP, Reichrath E, Verdonshot MM, Wijnhuizen GJ, Perenboom RJM. Development and validation of IMPACT-S, an ICF-based questionnaire to measure activities and participation. *J Rehabil Med* 2008; 40: 620–627.
11. Van Brakel WH, Anderson AM, Mutatkar RK, Bakirtzief Z, Nicholls PG, Raju MS, et al. The Participation scale: measuring a key concept in public health. *Disabil Rehabil* 2006; 28: 193–203.
12. Post MWM, de Witte LP. Good inter-rater reliability of the Frenchay Activities Index in stroke patients. *Clin Rehabil* 2003; 17: 548–552.
13. Van Brakel WH. *Participation Scale Users Manual Version 5.0*. Amsterdam; KIT Leprosy Unit; January 10, 2008.
14. Post WMW, van de Port I, Kap B, Berdenis van Berlekom S. Development and validation of the Utrecht Scale for Evaluation of Clinical Rehabilitation (USER). *Clin Rehabil* 2009; 23: 909–917.
15. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60: 34–42.
16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174.
17. Kramer MS, Feinstein AR. Clinical biostatistics. LIV. The biostatistics of concordance. *Clin Pharmacol Therapeutics* 1981; 29: 111–123.
18. Van Baalen B, Odding E, van Woensel MPC, van Kessel MA, Roebroek ME, Stam HJ. Reliability and sensitivity to change of measurement instruments used in a traumatic brain injury population. *Clin Rehabil* 2006; 20: 686–700.
19. Post MWM, Festen H, van de Port IG, Visser-Meily JMA. Reproducibility of the Caregiver Strain Index and the Caregiver Reac-

- tion Assessment in partners of stroke patients living in the Dutch community. *Clin Rehabil* 2007; 21:1050–1055.
20. Brown M, Dijkers MPJM, Gordon WA, Ashman T, Charatz H, Cheng Z. Participation Objective, Participation Subjective. *J Head Trauma Rehabil* 2004; 19: 459–481.
 21. Lemmens J, van Engelen EISM, Post WMW, Beurskens AJHM, Wolters PMJC, de Witte LP. Reproducibility and validity of the Dutch Life Habits Questionnaire (LIFE-H 3.0) in older adults. *Clin Rehabil* 2007; 21: 853–862.
 22. Lau ALD, Cummins RA. Test-retest reliability of the Personal Wellbeing Index [monograph on the Internet]. Hong Kong: The Hong Kong Polytechnic University; 2005 [cited 2009 July 16]. Available from: <http://acqol.deakin.edu.au/instruments/PWI/English-PWI-Adult.doc>