ORIGINAL REPORT

COMPARING THE RELIABILITY OF FIVE PARTICIPATION INSTRUMENTS IN PERSONS WITH SPINAL CONDITIONS

Vanessa K. Noonan, PhD PT[1,2], Jacek A. Kopec, MD PhD[2,3], Luc Noreau, PhD[4,5], Joel Singer, PhD[2,6], Louise C. Mâsse, PhD[7] and Marcel F. Dvorak, MD[1]

From the [1]Division of Spine, Department of Orthopaedics, [2]School of Population and Public Health, University of British Columbia, [3]Arthritis Research Centre of Canada, Vancouver, BC, [4]Rehabilitation Department, Laval University, [5]Centre for Interdisciplinary Research in Rehabilitation and Social Integration, Québec City, OC, [6]Canadian HIV Trials Network, and [7]Department of Pediatrics, University of British Columbia, Vancouver, BC, Canada

Objective: To compare the score distribution and reliability of 5 participation instruments developed using the International Classification of Functioning, Disability and Health.
Methods: Individuals treated for spinal conditions at an acute hospital were followed-up and 545 participated. Subjects completed 5 participation instruments (Impact on Participation and Autonomy (IPA), Keele Assessment of Participation (KAP), Participation Measure-Post Acute Care (PM-PAC), Participation Objective Participation Subjective (POPS) and World Health Organization Disability Assessment Schedule II (WHODAS II)). Test-retest reliability was assessed in 139 subjects. The score distribution, internal consistency and test-retest reliability were assessed.
Results: All the instruments demonstrated considerable ceiling effects, except for the POPS. Internal consistency (Cronbach's alpha) was ≥ 0.70 for all domains. The IPA and WHODAS II had the highest test-retest values, with intraclass correlation coefficients ≥ 0.70. The minimal detectable change as a percentage of the absolute scale score range was primarily between 20% and 30%.
Conclusion: The IPA, PM-PAC and WHODAS II have similar measurement properties. The KAP was designed for population-based studies and the POPS includes objective and subjective information, which may explain some of the differences observed. Researchers and clinicians should select an instrument that will fulfil their measurement objectives and future studies should assess minimal important change.

Key words: consumer participation; World Health Organization; rehabilitation; questionnaires; reproducibility of results.

INTRODUCTION

As disability rates continue to rise with an aging population and advances in medicine, there will be a greater need to understand how health conditions impact a person's life. The concept of participation, defined as the involvement in life situations in the International Classification of Functioning Disability and Health (ICF) (1), is receiving considerable attention. Since the ICF was published in draft form in 1997 a review conducted in 2008 identified 11 new participation instruments developed using the ICF (2).

Participation instruments can be used to assess individual and/or group differences, with individual differences being most important to clinicians, whereas group differences seem to be of greater interest to researchers (3). Criteria have been developed to help evaluate the measurement properties of instruments (4–6) aimed at detecting either individual or group differences, and two important criteria include assessing the score distribution (floor and ceiling effects) and reliability.

Floor and ceiling effects limit an instrument's ability to detect changes or differences in individuals or between groups (4). Ceiling effects have been reported in various participation instruments. One study reported that 70.3% of individuals with conditions such as diabetes reported no problems with self-care (7) and another study stated that 53% of community-dwelling individuals had no participation restrictions (including self-care) (8). Ceiling effects appear to be common in measuring participation, but very few studies have assessed ceiling effects and these instruments have not been directly compared (2).

Reliability is the degree to which the data produced by an instrument is free from random error (5). Two types of reliability are frequently assessed. The first is determining if the questions within a multi-item scale are homogenous or are internally consistent. The second type of reliability assesses whether the information provided by individuals remains stable over time (test-retest reliability in self-administered instruments). Intraclass correlation coefficients (ICC) are often used to report test-retest reliability and describe an instrument's ability to differentiate among individuals in the sample studied (9, 10). The variability of two measurements on the same individual can also be used to calculate the absolute measurement error called the standard error of measurement (SEM) (9, 10). There has been a growing interest in using the SEM to calculate the minimal amount of change in a score that must be observed given the absolute measurement error, referred to as the minimal detectable change (MDC) to determine if instruments can

detect individual changes in clinical practice (6). Many of these instruments have only recently been published and more testing is needed (2). It is also difficult to compare instruments, since the results are based on different health conditions and there are differences in how the instruments are administered in the studies.

A direct comparison of participation instruments would enable the provision of recommendations for clinicians and researchers. Persons with spinal conditions are an ideal population to evaluate participation instruments, since these conditions are prevalent and cause tremendous disability. Low back pain will affect 1 in 5 adults (11) and is reported to cost $100 billion per year in the USA, primarily due to an inability to work (11, 12). Spinal cord injuries (SCI) typically occur in males in their 30s, who will live a normal lifespan with their disability, and persons with SCI report severe limitations in self-care, recreation, fulfilling their family role and education (13). Finally, with an aging population there is an increase in spinal injuries from falls in the elderly causing spinal column fractures that affect all aspects of participation including self-care, mobility and community life (14, 15).

The purpose of this study was to compare the floor and ceiling effects, internal consistency, test-retest reliability (using ICCs and SEM) and MDC in 5 participation instruments in persons with spinal conditions. The 5 instruments include Impact on Participation and Autonomy (IPA) (16), Keele Assessment of Participation (KAP) (8), Participation Measure-Post Acute Care (PM-PAC) (17), Participation Objective and Participation Subjective (POPS) (18), and World Health Organization Disability Assessment Schedule II (WHODAS II) (19).

METHODS

Recruitment and study procedures

Adults admitted to the Vancouver General Hospital Spine Program between 2000 and 2005, were eligible if they had a diagnosis of: (*i*) a traumatic or non-traumatic spinal cord injury; (*ii*) a spinal column fracture without neurological involvement; or (*iii*) a spinal degenerative disease (e.g. disc herniation, spondylosis). A hospital database was used to identify subjects. Individuals were excluded if they were deceased; could not be contacted; did not speak English; had a cognitive deficit; were not able to physically complete the instruments (e.g. ventilator-dependent); or were discharged from hospital in the past 3 months and unable to perform regular activities (e.g. bed rest due to a pressure sore). These eligibility criteria were initially used to identify potential subjects, and the subjects' eligibility was re-assessed during the recruitment phase. A sample size of approximately 200 individuals with completed questionnaires was targeted for each spinal group for the cross-sectional study. Eligible individuals were randomly selected from the database until the target sample size was achieved or until all subjects were contacted. The study was approved by the Research Ethics Board at the University of British Columbia and all individuals provided written informed consent.

Individuals were contacted by mail and asked to complete a questionnaire. Test-retest reliability was assessed using a sub-set of subjects from the larger cross-sectional study who were asked to complete the instruments twice within 10 days. A target sample size of 50 individuals per diagnostic group was based on a sample size estimation of 124 subjects, which used an ICC of 0.75 obtained from the previous studies using these instruments (2). The sample size was estimated using a 95% confidence interval estimated for ICCs (20).

Data elements

Data was obtained from hospital databases and a questionnaire. The following types of data were collected: sociodemographic data (e.g. age, gender); socioeconomic data (e.g. education, employment); clinical data (e.g. diagnosis, neurological impairment, comorbidities); and scores from participation instruments. Neurological impairment was assessed in persons with traumatic SCI using the International Standards for the Neurological Classification of SCI (ISNCSCI) (21). Comorbidities at the time of follow-up were assessed using one section of the Self-Administered Comorbidity Questionnaire (22), which measures the presence or absence of 14 comorbid conditions (maximum score of 14).

Participation instruments

The IPA (16) assesses the perceived impact of a health condition or disability on participation and autonomy in the domains Autonomy Indoors (e.g. self-care); Family Role (e.g. housework); Autonomy Outdoors (e.g. visiting friends, leisure time); Social Life and Relationships; and Work and Education. The perceived participation score was calculated for each domain (*n* = 31 questions total), with a lower score indicating better perceived participation.

The KAP (8) contains 11 questions asking about autonomy in conducting life activities in the sub-domains Mobility; Self-Care; Domestic Life; Interpersonal Interactions and Relationships; Major Life Areas; and Community, Social and Civic Life. The mean score for each question in the KAP was compared with similar domains within the participation instruments. A lower score on a question in the KAP indicates better perceived participation.

The PM-PAC (17) is designed to assess participation in the community. It contains a total of 51 questions, and 42 questions are used to create scores for the domains Communication; Mobility; Domestic Life; Interpersonal Relationships; Role Functioning; Work and Employment; Education; Economic Life; and Community, Social and Civic Life. A higher score indicates better participation.

The POPS (18) assesses participation in 26 life activities from an objective (frequency) and subjective (importance and level of satisfaction) perspective. A scoring algorithm provided by the developers was used to calculate objective and subjective overall and domain scores (Domestic Life; Major Life Areas; Transportation; Interpersonal Interactions and Relationships; and Community, Recreational and Civic Life). Objective scores are based on z scores which represent the difference between the frequency information for each question compared with reference data from a sample of persons with traumatic brain injury (TBI) and healthy controls. The domains were weighted based on the perceived importance of the activity in the reference sample. Subjective scores are obtained by multiplying the individual's importance score by the satisfaction score. Subjective scores in the POPS range from −4 (important area that a person wants to do more or less of the activity) to +4 (important area that a person is satisfied with the amount of activity). The POPS was developed to be interviewer-administered and a self-administered version was tested for use in this study. The scoring algorithm was modified slightly when the raw (non-imputed) data was used. For this study we calculated domain scores even if less than half of the subjective questions were scored using the algorithm since the response "don't know" was considered missing data and it has been reported that if subjects do not engage in a specific activity (e.g. education) then they frequently omit additional questions asking if that particular activity is important to them (23). In addition, for the POPS objective domains the z-scores were not set at a maximum of −3 to +3 and instead the reported values were used.

The WHODAS II (19) assesses daily functioning using domains covered in the activities and participation component of the ICF and therefore measures both the concepts of activity and participation. There are 36 questions, and the domains include understanding and Communicating, Getting Around, Self-Care, Getting Along with People, Life Activities (household/work activities), and Participation in Society. A scoring algorithm was provided by the World Health Organization. Separate scores were calculated for individuals who

were working and not working for the Life Activities domain as well as for the total score. A lower score indicates better reported activity and participation.

Six participation instruments identified in the review of participation instruments (2) were excluded because they were: (*i*) too specific (Participation Survey/Mobility just measures lower extremity mobility(24)); (*ii*) designed to assess participation in developing countries (Participation Scale (25)); (*iii*) administered by interview (PAR-PRO (26)) or computer (Participation Measure-Post Acute Care-Computer Adaptive Test (27)), (*iv*) too similar to other instruments (Rating of Perceived Participation (28) was too similar to the IPA and KAP), or (*v*) not available (Perceived Impact of Problem Profile (29)).

Statistical analysis

For each instrument the score distribution, internal consistency and test-retest reliability (ICC and SEM) were evaluated. Score distributions were assessed using descriptive statistics (mean, standard deviation (SD), range). The percentage of individuals with the lowest level of participation and the highest level of participation were recorded and values greater than 15% were considered to be substantial floor and ceiling effects, respectively (6).

Internal consistency assesses the homogeneity of a multi-item scale, and it was evaluated using Cronbach's alpha coefficient (5). A minimum of 0.70 is recommended for group comparisons and 0.90–0.95 is needed for individual comparisons (5). In this study internal consistency was not assessed in the KAP or the POPS. The KAP was scored using individual questions, and even the overall score with the number of participation restrictions (each question dichotomized into yes or no) would likely not have high correlations among the questions. In the POPS, different aspects of participation are included in the domains, which are not necessarily related, and therefore measuring internal consistency is likely not applicable.

Test-retest reliability was assessed using a two-way random effects model (ICC2,1), with absolute agreement to account for any systematic variability between the two administrations. Recommended minimum values are 0.70 and 0.90 for individual and group level comparisons, respectively (4, 5). For instruments consisting of categorical scales, a weighted kappa coefficient was used. Although some suggest that it is difficult to apply a criterion when assessing weighted kappa (30), for the purpose of this study 0.70 was used as the minimal standard (6).

The SEM was calculated from the square root of the within-subject variance obtained from ANOVA (the square root of the sum of the between measures variance and the residual variance). Systematic differences between the test and retest were included when calculating the SEM, as recommended in the literature (10). The SEM can be used to calculate the MDC (MDC $= 1.96 \times \sqrt{2} \times$ SEM), which represents the smallest within-person change in score that can be detected in an individual beyond measurement error, with $\bar{p} < 0.05$ (6). The MDC as a percentage of the absolute theoretical scale score range was calculated to compare the instruments. Bland and Altman (31) recommend using plots to visually display the agreement between test administrations. The plots describe the differences between the first and second administration of the instrument against the average of the domain or total scores, and these plots were created for all of the instrument domains. Calculations for reliability, floor/ceiling, SEM and Bland and Altman plots were performed using SPSS 16.0 (Chicago, IL, USA).

Missing data was imputed using SAS 9.1.3 (Cary, NC, USA) program PROC MI and one simulated version of the data set was created for the overall sample (*n* = 545). Variables potentially related to the reason for the missing data or variables known to be associated with the participation scores were included in the model. The imputation was done within each instrument, and data pertaining to work and education were imputed only for individuals involved in these activities. Imputed data was used to estimate the score distribution and internal consistency, and non-imputed data was used to assess test-retest reliability (ICC, SEM estimates and Bland and Altman plots). The percentages of missing data for the questions in the first and second administration of the participation instruments was less than 10% (except for two ques-

tions in the second administration of the POPS asking about attending school, where the missing data was 12.8% and 14.1%).

RESULTS

A total of 545 individuals participated in the study (age range: 21–90 years). The response rates varied between 58% (187/320) in the spinal column group to 62% (213/345) in the spinal degenerative group. Average time of discharge from hospital to study follow-up was approximately 4 years [4.4 (2.2) years]. A total of 139 individuals completed the reliability study. The mean time and standard deviation between the first and second administration of the instruments was 14.7 (5.6) days and ranged between 7 and 31 days.

Table I. *Characteristics of the study respondents for the entire sample*

Variable	Overall (*n* = 545)	SCI (*n* = 145)	Spinal column (*n* = 187)	Spinal degenerative (*n* = 213)
Gender, %				
Male	67	79	71	56
Marital status, %				
Single	20	31	25	8
Married/partner	62	55	60	69
Divorced/widowed	18	14	15	23
Racial background, %				
Caucasian	86	80	88	87
Living support, %				
Live with someone	78	75	79	79
Education, %				
High school	39	43	36	38
College/university	49	49	54	45
Graduate degree	12	8	10	16
Employment, %				
Employed	52	32	70	50
Unemployed	3	7	2	2
Volunteer/retired	28	32	19	32
Unable to work	15	26	9	14
Compensation, %				
Yes	29	59	17	19
Spinal procedures, %				
Yes	78	86	48	98
AIS*, %				
AIS A		42		
AIS B		15		
AIS C		18		
AIS D		24		
Age, mean (SD), years				
At follow-up	51.5 (16.6)	48.7 (17.4)	46.8 (16.2)	57.6 (14.5)
Range	(21–90)	(21–86)	(21–85)	(24–90)
Comorbidity (0–14), mean (SD)				
Score at follow-up	1.2 (1.4)	1.0 (1.4)	0.9 (1.3)	1.5 (1.5)
Range	(0–8)	(0–8)	(0–6)	(0–7)
Motor score* (0–100), mean (SD)				
On admission		51.9 (26.2)		
Range		0–96		

*Subjects with traumatic spinal cord injury only (*n* = 123).
AIS: ASIA Impairment Scale; SCI: spinal cord injury; SD: standard deviation.

A description of the study subjects is provided in Table I. There were some notable differences in the 3 spine conditions. Sixty-seven percent ($n = 367$) were males and there were fewer males in the degenerative group (56%) compared with the SCI group (79%). There were differences in employment status, with the SCI group having the highest unemployment (7%) compared with the spinal column group and spinal degenerative group (2%). A comparison of individuals who participated in this study and those who were eligible, but did not participate revealed that the study participants were older (47.0 vs 40.0 years) on admission to hospital and there were fewer men (67% vs 73%) compared with non-participants.

Floor and ceiling effects

Scores for the instruments are reported in Table II. Ceiling effects were present in the IPA, KAP, PM-PAC and WHODAS II. The KAP had the highest percentages of ceiling effects, ranging from 56.7% to 75.8% in the 11 questions. All the IPA domains demonstrated ceiling effects ranging from 29.4% to 49.5%. Both the PM-PAC and the WHODAS II had some domains without ceiling effects; the PM-PAC domain Community, Social and Civic Life had a perfect score in 15% of the sample, and the 13.6% of subjects who did not work had a perfect score for the WHODAS II Life Activities domain. The POPS was the only instrument that did not have ceiling effects. In the POPS objective domains a numerical estimate of frequency is recorded (except the Domestic Life domain) and the questions are open-ended, making ceiling effects impossible. A floor effect was noted in the POPS objective Major Life Areas domain. None of the other instruments demonstrated any floor effects.

Internal consistency and test-retest reliability

Internal consistency was assessed using Cronbach's alpha (Table III). The internal consistency was good (values ≥ 0.70) in all the instruments and the IPA was the only instrument to have values for internal consistency ≥ 0.90. Internal consistency was not assessed in the KAP and the POPS.

In comparing the results from the first and second administration, most domain scores were not significantly different and there was no consistency in the direction of the change. The ICC values (95% confidence intervals) for the 5 instruments are reported in Table III. A comparison of the test-retest data for the POPS subjective domains using the original and the slightly modified scoring algorithm (generating a domain score even if less than half of the questions were scored) revealed that there was no impact on the ICCs and more subjects were included in the analysis for Transportation and Major Life Areas.

Estimates for the SEM and the MDC for each of the 5 instruments are summarized in Table III. The MDC as a percentage of the absolute scale range primarily ranged between 20% and 30%. The estimate of 13.5% for the Education domain in the PM-PAC and 67.0% for the Education question in the KAP were based on small samples. Due to the high ceiling effects it would not be possible to detect improvements beyond measurement error for most of the instruments (IPA,

KAP, PM-PAC, WHODAS II). It also would not be possible to detect deterioration in 3 POPS objective domains (Major Life Areas; Interpersonal Interactions and Relationships; Community, Recreational and Civic Life) due to floor effects. The Bland and Altman plots for the 5 instruments were reviewed and the differences between the 2 tests for each domain were not dependent on the levels of the domain scores in the 5 instruments (data not shown). A summary of the results is included in Table IV.

DISCUSSION

To our knowledge this is the first direct comparison of participation instruments that are based on the ICF. Overall, internal consistency estimates for the instruments' domains were acceptable (≥ 0.70); however, large ceiling effects were present in most of the instruments. The test-retest reliability data suggest that the instruments are able to discriminate at a group level (≥ 0.70). Estimates of the SEM and MDC indicate it would be difficult to detect improvements at an individual level due to the ceiling effects.

The results for ceiling effects are consistent with previous studies (7, 8). We observed large ceiling effects in domains related to self-care, economic life, and interpersonal interactions and relationships (IPA, KAP, PM-PAC, WHODAS II). Domains related to work as well as community, social and civic life had the least problems, but the percentage of the sample with a perfect score was still greater than 15% for most instruments. The KAP had the largest number of ceiling effects; 56% of the sample had a perfect score for each question. The IPA domains Autonomy Indoors as well as Social Life and Relationships are considered the least difficult, which is consistent with other studies (32–35). Ceiling effects were not an issue in the POPS because of scoring algorithm. For the POPS subjective domains, ceiling and floor effects were not common because it was rare for individuals to be completely unsatisfied or satisfied in all important areas for all questions within a domain. Ceiling effects observed in this study may result from subjects either recovering or adapting to their spinal condition and thereby not having many participation restrictions. If the purpose of collecting information about participation is to assess the effectiveness of an intervention, then it would be important to determine if the instruments are sensitive enough to detect a change as a result of an intervention. If, however, the intent is to determine if the subjects are able to achieve a criterion of being acceptable, then the issue of the ceiling effects observed in this study may not be a problem.

Estimates of internal consistency were very good in the IPA, PM-PAC and WHODAS II. All of the IPA domains had a Cronbach's alpha values between 0.90 and 0.96, and IPA was the only instrument that met the criteria for both individual and group comparisons. The IPA Social Life and Relationships domain had the lowest value for internal consistency (0.90), and this is supported by other studies (16, 35, 36). Internal consistency was also lowest in the domain Getting Along with People (0.81) in the WHODAS II and other studies assessing individuals with health

Table II. *Descriptive information and floor/ceiling effects for the first administration of the participation instruments based on the entire sample* (n = *545*)

Instruments (score range)	Overall mean (SD)	Overall range	% Worst possible score	% Best possible score
IPA (0–4)				
Autonomy Indoors	0.55 (0.77)	0–3.57	0.0	49.5
Family Role	0.99 (0.97)	0–4.00	0.2	29.4
Autonomy Outdoors	1.14 (1.14)	0–4.00	1.5	31.0
Social Life & Relationships	0.62 (0.70)	0–3.00	0.0	41.1
Work & Education (*n*=356)	0.99 (1.12)	0–4.00	1.7	38.2
KAP (1–5)				
Mobility #1	1.40 (0.73)	1.00–5.00	0.4	70.3
Mobility #2	1.69 (0.97)	1.00–5.00	1.5	56.7
Self-Care	1.37 (0.78)	1.00–5.00	1.1	75.8
Domestic Life #4	1.62 (0.95)	1.00–5.00	1.8	61.5
Domestic Life #5	1.45 (0.81)	1.00–5.00	1.1	69.5
Domestic Life #6 (*n*=286)	1.58 (0.87)	1.00–5.00	1.4	60.1
Interpersonal Interactions & Relationships	1.49 (0.82)	1.00–5.00	0.9	66.6
Economic Life	1.48 (1.00)	1.00–5.00	5.7	74.7
Work (*n*=327)	1.57 (1.10)	1.00–5.00	5.8	71.6
Education (*n*=193)	2.05 (1.48)	1.00–5.00	14.0	58.0
Community, Social & Civic Life (*n*=412)	1.70 (1.08)	1.00–5.00	3.6	60.9
PM-PAC (1–5)				
Communication	4.63 (0.66)	1.00–5.00	0.4	58.2
Mobility	4.26 (0.93)	1.00–5.00	0.2	43.3
Domestic Life	4.32 (0.87)	1.00–5.00	0.6	44.8
Interpersonal Relationships	4.08 (0.94)	1.00–5.00	0.4	30.8
Role Functioning	3.54 (1.19)	1.00–5.00	4.0	16.7
Work & Employment (*n*=299)	4.19 (0.97)	1.00–5.00	1.0	39.1
Education (*n*=63)	4.43 (0.78)	2.00–5.00	0.0	43.8
Economic Life	4.59 (0.76)	1.00–5.00	0.6	66.6
Community, Social & Civic Life	4.03 (0.90)	1.17–5.00	0.0	15.0
POPS: Objective*				
Objective Domestic Life	–0.15 (0.91)	–2.22–2.03	0.6	2.0
Objective Major Life Areas	0.79 (1.76)	–0.98–10.69	27.5	NA
Objective Transportation	–0.80 (0.56)	–1.31–3.17	2.0	NA
Objective Interpersonal Interactions & Relationships	0.88 (2.54)	–1.59–20.09	0.7	NA
Objective Community, Recreational & Civic Life	0.43 (1.37)	–1.16–10.06	1.1	NA
Objective Participation Total	0.24 (0.91)	–1.29–4.34	0.0	NA
POPS: Subjective (–4 to 4)				
Subjective Domestic Life	1.00 (1.28)	–3.00–4.00	0.0	0.4
Subjective Major Life Areas	0.28 (1.44)	–3.33–3.33	0.0	0.0
Subjective Transportation	0.89 (1.41)	–4.00–4.00	0.2	0.6
Subjective Interpersonal Interactions & Relationships	0.99 (1.19)	–3.38–3.75	0.0	0.0
Subjective Community, Recreational & Civic Life	0.70 (0.96)	–2.80–3.20	0.0	0.0
Subjective Participation Total	0.77 (0.88)	–2.77–2.92	0.0	0.0
WHODAS II (0–100)				
Understanding & Communicating	11.48 (16.69)	0–80.00	0.0	48.1
Getting Around	31.33 (27.57)	0–100.00	1.3	22.4
Self-Care	13.74 (22.20)	0–100.00	0.9	61.0
Life Activities (Non-working; *n*=162)	45.56 (30.95)	0–100.00	10.5	13.6
Life Activities (Working; *n*=383)	21.64 (23.93)	0–100.00	1.0	33.2
Getting Along with People	16.07 (19.79)	0–100.00	0.2	40.2
Participation in Society	26.93 (22.43)	0–91.67	0.0	17.4
Total Score (Non-working; *n*=162)	29.91 (17.26)	0–76.09	0.0	2.5
Total Score (Working; *n*=383)	18.20 (17.58)	0–84.91	0.0	12.5

*The score range for the POPS objective domains varies.

IPA: Impact on Participation and Autonomy; KAP: Keele Assessment of Participation; NA: Not applicable; PM-PAC: Participation Measure-Post Acute Care; POPS: Participation Objective Participation Subjective; SD: standard deviation; WHODAS II: World Health Organization Disability Assessment Schedule II.

Table III. *Internal consistency, test-retest reliability and standard error of measurement*

Instruments (score range)	# Questions	Cronbach's alpha (*n*=545)	ICC (95% CI) Weighted kappa (95% CI)* (*n*=139)	SEM	MDC	MDC %
IPA (0–4)						
Autonomy Indoors	7	0.94	0.84 (0.78, 0.88)	0.25	0.70	17.5
Family Role	7	0.95	0.88 (0.84, 0.92)	0.30	0.83	20.8
Autonomy Outdoors	5	0.95	0.85 (0.80, 0.89)	0.42	1.18	29.0
Social Life & Relationships	6	0.90	0.83 (0.77, 0.88)	0.28	0.76	19.0
Work and Education (*n* = 71)	6	0.96	0.86 (0.79, 0.91)	0.35	0.96	24.0
KAP (1–5)*						
Mobility #1	1	NA	0.60 (0.45, 0.76)	0.31	0.88	22.0
Mobility #2	1	NA	0.61 (0.49, 0.73)	0.54	1.05	26.3
Self-Care	1	NA	0.54 (0.40, 0.68)	0.33	0.91	22.8
Domestic Life #4	1	NA	0.61 (0.49, 0.73)	0.39	1.09	27.3
Domestic Life #5	1	NA	0.49 (0.35, 0.63)	0.40	1.10	27.6
Domestic Life #6	1	NA	0.79 (0.66, 0.92)	0.26	0.72	18.0
Interpersonal Interactions & Relationships	1	NA	0.65 (0.53, 0.77)	0.33	0.91	22.8
Economic Life	1	NA	0.47 (0.27, 0.67)	0.62	1.73	43.3
Work (*n* = 75)	1	NA	0.67 (0.49, 0.85)	0.48	1.34	33.5
Education (*n* = 39)	1	NA	0.56 (0.32, 0.80)	0.97	2.68	67.0
Community, Social & Civic Life (*n* = 101)	1	NA	0.67 (0.55, 0.79)	0.40	1.10	27.5
PM-PAC (1–5)						
Communication	6	0.91	0.59 (0.47, 0.69)	0.29	0.80	20.0
Mobility	5	0.93	0.91 (0.87, 0.93)	0.26	0.73	18.3
Domestic Life	3	0.85	0.81 (0.74, 0.86)	0.34	0.94	23.5
Interpersonal Relationships	3	0.85	0.76 (0.68, 0.82)	0.42	1.17	29.3
Role Functioning	4	0.92	0.74 (0.65, 0.81)	0.58	1.61	40.3
Work & Employment (*n* = 69)	5	0.90	0.78 (0.66, 0.86)	0.42	1.16	29.0
Education (*n* = 13)	4	0.84	0.88 (0.65, 0.96)	0.19	0.54	13.5
Economic Life	3	0.84	0.77 (0.69, 0.83)	0.30	0.84	21.0
Community, Social & Civic Life	9	0.90	0.83 (0.77, 0.88)	0.34	0.93	23.3
POPS: Objective†						
Objective Domestic Life	8	NA	0.90 (0.87, 0.93)	0.28	0.79	NA
Objective Major Life Areas	3	NA	0.86 (0.81, 0.90)	0.56	1.54	NA
Objective Transportation	2	NA	0.78 (0.71, 0.84)	0.23	0.64	NA
Objective Interpersonal Interactions & Relationships	8	NA	0.61 (0.49, 0.70)	1.20	3.33	NA
Objective Community, Recreational & Civic Life	5	NA	0.66 (0.55, 0.74)	0.95	2.62	NA
Objective Participation Total	26	NA	0.82 (0.75, 0.87)	0.34	0.93	NA
POPS: Subjective (–4 to 4)						
Subjective Domestic Life	16	NA	0.67 (0.57, 0.75)	0.70	1.93	24.1
Subjective Major Life Areas	6	NA	0.64 (0.53, 0.74)	1.03	2.86	35.8
Subjective Transportation	4	NA	0.63 (0.51, 0.72)	0.99	2.74	34.3
Subjective Interpersonal Interactions & Relationships	16	NA	0.72 (0.63, 0.79)	0.65	1.81	22.6
Subjective Community, Recreational & Civic Life	10	NA	0.61 (0.49, 0.70)	0.67	1.86	23.2
Subjective Participation Total	52	NA	0.82 (0.76, 0.93)	0.43	1.19	14.9
WHODAS II (0–100)						
Understanding & Communicating	6	0.90	0.79 (0.72, 0.85)	6.37	17.64	17.6
Getting Around	5	0.85	0.90 (0.87, 0.93)	8.24	22.82	22.8
Self-Care	4	0.85	0.87 (0.83, 0.91)	6.18	17.12	17.1
Life Activities (Non-working) (*n* = 53)	4	0.91	0.74 (0.59, 0.84)	15.44	42.77	42.8
Life Activities (Working) (*n* = 86)	8	0.94	0.87 (0.81, 0.92)	7.45	20.64	20.6
Getting Along with People	5	0.81	0.72 (0.63, 0.79)	9.04	25.04	25.0
Participation in Society	8	0.90	0.85 (0.78, 0.89)	8.01	20.64	20.6
Total Score (Non-working) (*n* = 53)	32	0.94	0.89 (0.82, 0.94)	5.51	15.26	15.3
Total Score (Working) (*n* = 86)	36	0.96	0.91 (0.86, 0.94)	4.69	12.99	13.0

*Test-retest reliability was assessed using weighed kappa for the KAP.

†The score range for the POPS objective domains varies.

MDC: minimal detectable change; NA: not applicable; SEM: standard error of measurement; see Table II for instrument abbreviations.

Table IV. *Summary* of the study results for the score distribution and reliability*

Criteria	IPA	KAP	PM-PAC	POPS OBJ	POPS SUBJ	WHODASII
Score Distribution (Floor/Ceiling)	+	+	+	+++	+++	++
Reliability						
1) Internal Consistency	+++	NA	+++	NA	NA	+++
2) Test-retest Reliability (ICC/weighted kappa)	++	+	++	++	+	++
3) Test-retest Reliability (SEM/MDC)	++	+	++	NA	++	++

*Ratings: +++ met criteria; ++ partially met criteria; + results primarily did not meet criteria.
ICC: intraclass correlation coefficient; MDC: minimal detectable change; NA: not applicable; SEM: standard error of measurement; see Table II for instrument abbreviations.

conditions such as stroke, breast cancer, diabetes and osteoarthritis have reported similar findings (7). Internal consistency was not assessed in the KAP and the POPS. It has been suggested that not all the measurement criteria are necessarily relevant when assessing instruments (37). The POPS domains include different types of information, for example there are questions about school and work in one domain, and it does not make sense to assume that a person who works will also attend school (18). In the IPA and WHODAS II there was one domain covering work and school; however, the subjects select the activity that is most relevant to them when answering the questions that will make the responses to the questions more homogenous.

Test-retest reliability estimates based on ICC values were adequate in the 5 participation instruments and our results are similar to other studies. Domains in the IPA, WHODAS II and the PM-PAC (except the Communication domain) met the criterion of having an ICC ≥ 0.70 suggested for measuring group differences and are similar to other studies (16, 17, 38, 39). Very few studies have demonstrated that these participation instruments are able to achieve ICC values ≥ 0.90 recommended for individual comparisons. The study by Sibley et al. (35) was one of the few studies that reported ICC values for the IPA between 0.91 and 0.97, with a two-week interval between tests. Test-retest reliability was higher for the objective participation domains compared with the subjective participation domains in the POPS, which was not the case in individuals with TBI (18). Differences in the ICC values may be due to sample variability or type of ICC used.

Results from this study add new information regarding the absolute measurement error. Estimates of SEM and the MDC have been previously reported for the WHODAS II in adults with acquired hearing loss, and estimates were higher in our study for 5 of the 7 domains (38). There are variations in the type of data and calculations used to calculate SEM and MDC (10) (e.g. SEM can be calculated using Cronbach's alpha or within-subject variance from test-retest studies), and the methods used by Chisolm et al. (38) were not explicitly described, which may explain the differences observed.

The values for the MDC as a percentage of the absolute scale score mostly ranged between 20% and 30% (range 13.5–67%) for the participation domains. Other studies have also reported values ranging between 26% and 39% for instruments such as the Low Vision Quality of Life and Vision-Related Quality of Life Core Measure (40). For the Sickness Impact Profile measurement error accounted for 9.3% in the total score and was as high as 40.3% for questions asking about alertness, suggesting that it was not responsive enough to detect changes in individuals who had a stroke (9). In this study the high ceiling effects would make it impossible to detect improvements beyond measurement error for the majority of domains in the IPA, KAP, PM-PAC and WHODAS II in these diagnostic groups, which may be because these instruments were developed using other health conditions. The MDC estimates are based on individual level changes, and group level MDC estimates would be lower since the MDC is divided by the square root of the sample size, making it easier to detect group changes (41). The MDCs do not necessarily represent the differences that are expected to be clinically relevant, which are referred to as minimal important change (MIC). Future studies must further assess changes that are meaningful to individuals receiving a particular intervention, and for an instrument to be clinically useful the MDC should be less than the MIC (9).

There are several limitations to this study. Although the sample included 3 spinal conditions with differing clinical symptoms, these results cannot be generalized to other health conditions. Future studies should continue to compare the instruments in persons with more disabling health conditions given the problems experienced with ceiling effects. This study was a cross-sectional assessment of participation following an acute care admission and future studies should establish the MIC in longitudinal studies before any conclusions can be drawn regarding their role in clinical assessment.

In conclusion, this study compared the score distributions, internal consistency and test-retest reliability for 5 participation instruments. The IPA, PM-PAC and WHODAS II had similar measurement properties. The KAP was developed to assess participation in population-based studies and our results suggest it is likely not suitable for clinical practice. The POPS is unique as it captures both objective and subjective information. Measuring changes in participation at the individual level with the current instruments may be difficult due to measurement error. Researchers and clinicians should consider the type of information they require and the measurement properties before selecting an instrument.

REFERENCES

1. World Health Organization. International Classification of Functioning, Disability and Health. Geneva (Switzerland): World Health Organization; 2001.
2. Noonan VK, Kopec JA, Noreau L, Singer J, Dvorak MF. A review of participation instruments based on the International Classification of Functioning, Disability and Health. Disabil Rehabil 2009; 31: 1883–1901.
3. Dekker J, Dallmeijer AJ, Lankhorst GJ. Clinimetrics in rehabilitation medicine: current issues in developing and applying measurement instruments. J Rehabil Med 2005; 37: 193–201.
4. 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.
5. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res 2002; 11: 193–205.
6. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007; 60: 34–42.
7. Posl M. Evaluation of the World Health Organization Disability Assessment Schedule II (WHODAS II) – German Version [dissertation]. Ludwig Maximilian University: Munich (Germany); 2004.
8. Wilkie R, Peat G, Thomas E, Hooper H, Croft PR. The Keele Assessment of Participation: a new instrument to measure participation restriction in population studies. Combined qualitative and quantitative examination of its psychometric properties. Qual Life Res 2005; 14: 1889–1899.
9. Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek AL. Smallest real difference, a link between reproducibility and responsiveness. Qual Life Res 2001; 10: 571–578.
10. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 2006; 59: 1033–1039.
11. Cassidy JD, Cote P, Carroll LJ, Kristman V. Incidence and course of low back pain episodes in the general population. Spine 2005; 30: 2817–2823.
12. Katz JN. Lumbar disc disorders and low-back pain: socioeconomic factors and consequences. J Bone Joint Surg Am 2006; 88 Suppl 2: 21–24.
13. Lund ML, Nordlund A, Nygard L, Lexell J, Bernspang B. Perceptions of participation and predictors of perceived problems with participation in persons with spinal cord injury. J Rehabil Med 2005; 37: 3–8.
14. Kannus P, Niemi S, Palvanen M, Parkkari J. Continuously increasing number and incidence of fall-induced, fracture-associated, spinal cord injuries in elderly persons. Arch Intern Med 2000; 160: 2145–2149.
15. Tinetti ME, Baker DI, McAvay G, Claus EB, Garrett P, Gottschalk M, et al. A multifactorial intervention to reduce the risk of falling among elderly people living in the community. N Engl J Med 1994; 331: 821–827.
16. Cardol M, de Haan RJ, de Jong BA, van den Bos GA, de Groot IJ. Psychometric properties of the Impact on Participation and Autonomy Questionnaire. Arch Phys Med Rehabil 2001; 82: 210–216.
17. Gandek B, Sinclair SJ, Jette AM, Ware JE, Jr. Development and

18. initial psychometric evaluation of the Participation Measure for Post-Acute Care (PM-PAC). Am J Phys Med Rehabil 2006; 86: 57–71.
18. Brown M, Dijkers MP, Gordon WA, Ashman T, Charatz H, Cheng Z. Participation objective, Participation subjective: a measure of participation combining outsider and insider perspectives. J Head Trauma Rehabil 2004; 19: 459–481.
19. World Health Organization. Geneva: WHODAS II [updated 2001, November 27; cited 2006 March 10]. Available from: http://www.who.int/icidh/whodas.
20. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 2nd edn. Oxford: Oxford University Press; 1995.
21. Marino RJ, Barros T, Biering-Sorensen F, Burns SP, Donovan WH, Graves DE, et al. International standards for neurological classification of spinal cord injury. J Spinal Cord Med 2003; 26 Suppl 1: S50–S56.
22. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Rheum 2003; 49: 156–163.
23. Dijkers M. Comments on van Brakel et al.'s Participation Scale. Disabil Rehabil 2006; 28: 1360–1362.
24. Gray DB, Hollingsworth HH, Stark SL, Morgan KA. Participation Survey/Mobility: psychometric properties of a measure of participation for people with mobility impairments and limitations. Arch Phys Med Rehabil 2006; 87: 189–197.
25. 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.
26. Ostir GV, Granger CV, Black T, Roberts P, Burgos L, Martinkewiz P, et al. Preliminary results for the PAR-PRO: a measure of home and community participation. Arch Phys Med Rehabil 2006; 87: 1043–1051.
27. Haley SM, Gandek B, Siebens H, Black-Schaffer RM, Sinclair SJ, Tao W, et al. Computerized adaptive testing for follow-up after discharge from inpatient rehabilitation: II. Participation outcomes. Arch Phys Med Rehabil 2008; 89: 275–283.
28. Sandstrom M, Lundin-Olsson L. Development and evaluation of a new questionnaire for rating perceived participation. Clin Rehabil 2007; 21: 833–845.
29. Pallant JF, Misajon R, Bennett E, Manderson L. Measuring the impact and distress of health problems from the individual's perspective: development of the Perceived Impact of Problem Profile (PIPP). Health Qual Life Outcomes 2006; 4: 36.
30. Fayers PM, Machin D. Quality of life: assessment, analysis and interpretation. Chichester: John Wiley and Sons; 2000.
31. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999; 8: 135–160.
32. Franchignoni F, Ferriero G, Giordano A, Guglielmi V, Picco D. Rasch psychometric validation of the Impact on Participation and Autonomy questionnaire in people with Parkinson's disease. Eura Medicophys 2007; 43: 451–461.
33. Lund ML, Fisher AG, Lexell J, Bernspang B. Impact on Participation and Autonomy questionnaire: internal scale validity of the Swedish version for use in people with spinal cord injury. J Rehabil Med 2007; 39: 156–162.
34. Middelkamp W, Moulaert VR, Verbunt JA, van Heugten CM, Bakx WG, Wade DT. Life after survival: long-term daily life functioning and quality of life of patients with hypoxic brain injury as a result of a cardiac arrest. Clin Rehabil 2007; 21: 425–431.
35. Sibley A, Kersten P, Ward CD, White B, Mehta R, George S. Measuring autonomy in disabled people: Validation of a new scale in a UK population. Clin Rehabil 2006; 20: 793–803.
36. van Genderen FR, van Meeteren NL, Heijnen L, van den Berg HM, Helders PJ. The use of a disability model in haemophilia research. Haemophilia 2005; 11: 472–480.
37. Dijkers MP, Whiteneck G, El Jaroudi R. Measures of social outcomes in disability research. Arch Phys Med Rehabil 2000;

81: S63–S80.
38. Chisolm TH, Abrams HB, McArdle R, Wilson RH, Doyle PJ. The WHO-DAS II: psychometric properties in the measurement of functional health status in adults with acquired hearing loss. Trends Amplif 2005; 9: 111–126.
39. McKibbin C, Patterson TL, Jeste DV. Assessing disability in older patients with schizophrenia: results from the WHODAS-II. J Nerv Ment Dis 2004; 192: 405–413.

40. de Boer MR, de Vet HC, Terwee CB, Moll AC, Volker-Dieben HJ, van Rens GH. Changes to the subscales of two vision-related quality of life questionnaires are proposed. J Clin Epidemiol 2005; 58: 1260–1268.
41. de Vet HC, Bouter LM, Bezemer PD, Beurskens AJ. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. Int J Technol Assess Health Care 2001; 17: 479–487.