Online Publication Date: 01 January 2007
To cite this Article: Kersten, Paula, Cardol, Mieke, George, Steve, Ward, Christopher, Sibley, Andrew and White, Barney (2007) 'Validity of the impact on participation and autonomy questionnaire: A comparison between two countries ',
Disability & Rehabilitation, 29:19, 1502 - 1509
To link to this article: DOI: 10.1080/09638280601030066
URL: http://dx.doi.org/10.1080/09638280601030066

Validity of the Impact on Participation and Autonomy questionnaire – a comparison between two countries

Paula Kersten, PhD[1] P.Kersten@soton.ac.uk

Mieke Cardol, PhD[2] m.cardol@nivel.nl

Steve George, MD[3] pluto@soton.ac.uk

Christopher Ward, MD[4] C.D.Ward@nottingham.ac.uk

Andrew Sibley, MSc[5] A.Sibley@soton.ac.uk

Barney White, MSc[3] barney.white@southbirminghampct.nhs.uk

1. School of Health Professions and Rehabilitation Sciences, University of Southampton, Southampton, United Kingdom.
2. NIVEL (Netherlands Institute for Health Services Research), Utrecht, the Netherlands & Department of Rehabilitation, Amsterdam Medical Centre, Amsterdam, The Netherlands.
3. Public Health Sciences, School of Medicine, University of Southampton, Southampton, United Kingdom.
4. Rehabilitation Research Unit, University of Nottingham, Nottingham, United Kingdom
5. School of Nursing & Midwifery, University of Southampton, Southampton, United Kingdom

The study was conducted in Hampshire, Wiltshire, Derbyshire and Nottinghamshire in the United Kingdom and in North Holland in the Netherlands.

All correspondence (including requests for reprints) to:

Paula Kersten, PhD

Senior Lecturer

School of Health Professions and Rehabilitation Sciences

University of Southampton

Highfield

Southampton

SO17 1BJ

United Kingdom

Tel +44 23 80595299

Fax +44 23 80595301

Email P.Kersten@soton.ac.uk

Running Title: Cross-cultural validity of the IPA

Keywords: Outcome Assessment (Health Care); Rehabilitation; Chronic Disease

The cross-cultural validity of the Impact on Participation and Autonomy questionnaire

Abstract

Purpose

To evaluate the cross-cultural validity of the five subscales of the Impact on Participation and Autonomy (IPA) measure and the full 31-item scale.

Method

Data from two validation studies (Dutch and English) were pooled (n=106). Participants (aged 18-75), known to rehabilitation services or GP practices, had conditions ranging from minor ailments to significant disability. Validity of the five subscales and the total scale was examined using Rasch analysis (Partial Credit Model). P-values smaller than 0.01 were employed to allow for multiple testing.

Results

A number of items in all the subscales except 'Outdoor Autonomy' needed rescoring. One 'Indoor Autonomy' item showed uniform DIF by country and was split by country. One 'Work and Education' item displayed uniform and non-uniform DIF by gender. All the subscales fitted the Rasch model and were invariant across country. A 30-item IPA also fitted the Rasch model.

Conclusion

The IPA subscales and a 30-item scale are invariant across the two cultures and gender. The IPA can be used validly to assess participation and autonomy in these

populations. Further analyses are required to examine whether the IPA is invariant across differing levels of disability and other disease groups not included in this study.

The cross-cultural validity of the Impact on Participation and Autonomy questionnaire

Introduction

On an individual level, the experiences of a chronic illness are numerous and complex. Medical factors may not be the most crucial for treatment, but rather personal context, experiences and needs define a patient's priorities and goals. Since rehabilitation is concerned with alleviation of the long-term consequences of disease, assessment should address long-term outcomes in terms of participation. The concept of autonomy adds a personal perspective to the assessment of participation.[1] For example, a person's participation can be demonstrated through life roles such as that of a worker, but further information is required in order to determine the extent of personal choice. Therefore, in the domain of participation an important question might be 'how much scope do individuals have for living their lives as they want? ' The *Impact on Participation and Autonomy* Questionnaire (IPA) is a relatively new generic outcome measure that evaluates the perceived personal impact of chronic disability on participation and autonomy.[2,3] The IPA was developed in the Netherlands in 2001 and validated with people with neuromuscular disease, spinal cord injury, stroke, rheumatoid arthritis, traumatic hand injury and fibromyalgia. The IPA was shown to be valid and reliable, with promising responsiveness properties.[2-6] The IPA is now used in clinical practice as well as in rehabilitation research, nationally and internationally.[7-9] The original Dutch IPA has 31 items, measuring participation and autonomy, that have been shown to load onto four factors, i.e. 'autonomy indoors'; 'family role'; 'autonomy outdoors'; and 'social life and relationships'. A fifth subscale, 'work and education', has been kept throughout the

Dutch validation studies of the IPA. However, as many of the Dutch participants in the initial studies were not employed, this subscale was not confirmed as an independent factor. In a subsequent English validation study the factor structure has been confirmed, including this fifth factor 'work and education'.[9] The number of items in each subscale varies (Box 1) and responses to each item range from zero to four with higher scores reflecting more (negative) impact on participation and autonomy. The IPA also contains eight questions that ask people to report on whether they perceive their limitations in participation as problematic. These items help to quantify the impact of disablement. The English version included also an additional item covering helping and supporting other people. This item had been added to the Dutch version after its publication. In an English validation study this item was shown to load onto the 'social life and relationships' subscale.[9]

Box 1 Subscales structure to be tested in the Rasch Analysis

IPA Subscales	**IPA Items**
1. Autonomy indoors (7 items)	1a, 1b, 2a, 2b, 2c, 2d, 2e
2. Family role (7 items)	3a, 3b, 3c, 3d, 3e, 3f, 4a
3. Autonomy outdoors (5 items)	1c, 1d, 5a, 6g, 10
4. Social life and relationships (7 items)	6a, 6b, 6c, 6d, 6e, 6f
5. Work and education (6 items)	7a, 7b, 7c, 7d, 7e, 8a

The IPA is increasingly being adopted internationally. Since autonomy and participation may be affected by cultural factors, the purpose of the present study was to investigate the cross-cultural validity of the IPA. The analysis was conducted solely on the original 31 participation and autonomy items, since we did not have Dutch data on the new item. Thus, the analysis aimed to find out whether the IPA evaluates participation and autonomy the same way in the Netherlands as it does in the United Kingdom.

Methods

The analysis presented here used data from two surveys, both of which have been described in detail before.[6,9] The Dutch study examined the responsiveness of the IPA and was conducted in an outpatient clinic of a rehabilitation department of an academic hospital. Fifty-three persons with various chronic conditions (e.g. stroke, neuromuscular disorder, severe hand injury) were enrolled in the study, 68% female, 32% male, median age 50 yrs. The study sample included competent Dutch speakers aged 18 to 75 years, who had recently been admitted for rehabilitation treatment. For the English study the IPA was adapted, using strict guidelines.[10] The sample included 213 competent English speakers aged 18 to 75 years (42% male, 58% female; median age 54), with multiple sclerosis, rheumatoid arthritis, spinal cord injury (recruited via out-patients) and minor ailments (recruited through GP practices). Both studies had been approved by relevant ethics committees and all participants gave informed consent.

Data analysis

We used Rasch analysis to test the cross-cultural validity of the IPA . The Rasch model is a unidimensional model which asserts that more able people (here in terms of participation and autonomy) are more likely to affirm an item compared to less able people and vice versa.[11,12] There are two Rasch models that can cope with polytomous data.[13] They are different in the way they deal with item thresholds: thresholds are the points where the probabilities of a response of either 0 or 1, and 1 or 2 (and so forth) are equally likely. The first model, the *Rasch Rating Scale Model*, assumes that, irrespective of the items having a different level of difficulty, the thresholds will have the same distance between them, i.e. the items share the same structure. By contrast, the *Partial Credit Model* makes no assumptions about the equality of the threshold locations relative to each item. To choose the correct model to be used in the analysis we conducted a log-likelihood test, which showed that there was a significant difference between the two models for our data ($\chi^2 = 4211.59$; $P < 0.001$). Therefore the *Partial Credit Model* (Equation 1) was used.

Equation 1

$$\ln\left(\frac{P_{nik}}{1-P_{nik-1}}\right) = \theta_n - b_{ik}$$

where P is the probability of person *n* affirming category *k* in item *i*; compared with an adjacent category *(k-1)*.

The analysis was conducted for each IPA subscale separately and will be described in further detail:

1. Threshold ordering: Rasch log-transformed item scores generated from the response choices should reflect the increasing or decreasing latent trait to be measured. For example, a person with a very low location along the trait (in the case of the IPA reflecting very good autonomy) relative to the item location should have a greater probability of selecting the response category 0. By contrast a person with a location much higher than the item location is most likely to select the response category 4. If the IPA categories indeed reflect increasing participation and autonomy, then we would expect thresholds defining the categories to be ordered along the trait likewise.[12] Optimal items are items where the thresholds are ordered. However, disordered thresholds can also be observed. This means that people with a given level of participation and autonomy do not affirm the expected response option to an item. Where this occurs it will be necessary to collapse item categories (i.e. group them together) until they are ordered. After this process (also called rescoring) the data are re-examined to establish the overall fit to the model and how well each item fits the model.
2. Fit to the model was also examined with summary fit statistics:
    - The item fit residual statistic and the person fit residual statistic are distributed as a Z-statistic with a mean of zero and a standard deviation (SD) of 1, which indicate perfect fit. Thus, z-statistics close to zero (with SD close to 1) would indicate the data fit the Rasch model;
    - The item-trait summary statistic, the χ^2 reflects the property of invariance across the trait and should therefore be insignificant;
    - The person separation index (PSI) is an indicator of how precisely subjects have been spread out along the measurement construct defined by the items.[14]

In other words, how well does the measure identify discrete groups of people? This value should be greater than 0.80.

3. Individual item fit residual statistics summate individual item deviations. They are deemed acceptable within the range of ± 2.5: a high negative residual suggests that an item is redundant and can be removed, whereas a high positive residual suggests an item does not fit the Rasch model and should be removed.
4. Differential Item Functioning (DIF) analysis: The hypothesis we were testing was that the IPA behaves in the same way in a Dutch and a UK sample.[15]. Therefore, the location of items along the measurement construct should be the same in the two samples. This was examined with Item Characteristics Curves (ICC's) and Analysis of Variance.[16] If the measurement construct under consideration (i.e. the subscale) is unidimensional and free of cultural bias, then (except for random variation) we should find that the Dutch and UK ICC's have the same shape and location.[17] Items that do not yield the same item response function for two or more groups display DIF and are violating the requirement of unidimensionality.[18] When items display a constant difference between groups in the probability of affirming an item category across the construct, the item is said to display *uniform DIF*. These items can be split by country. When the differences vary across the construct the item is displaying *non-uniform DIF*. Since it is not possible to adjust for non-uniform DIF, those items should be removed from the scale. DIF analysis was also conducted to examine bias by gender groups.
5. Principal Component Analysis (PCA) of the residuals (of the final subscale): the residuals are what remain when the 'Rasch Factor' has been removed from the data. Thus, the first factor of the Principal Components Analysis is the primary contributor to the variance of the data, with the 'Rasch Factor' discounted.[19]

Five analyses were conducted separately for each subscale. In addition, the entire analysis was repeated for all the data combined, discounting the proposed subscales, so as to test if the 31-items IPA would be a unidimensional scale. To conclude this 31-item scale analysis a formal test of the assumption of local independence was conducted[19], to investigate whether any subset of the items in the scale would measure the same dimension as the complete scale. We tested the possibility that patterns of items identified in the residuals might have an effect upon person estimates. For this purpose two paired t-tests were conducted comparing person locations that were estimated using two subsets of items taken from the final scale, and the final scale as a whole.

Throughout, P-values smaller than 0.01 were employed to allow for multiple testing.[20] All analyses were conducted in RUMM2020.[21]

Sample size

For Rasch analyses reasonably well targeted samples of 50 have 99% confidence that the estimated item difficulty is within +/- 1 logit of its stable value (especially when persons take 10 or more items).[22] Our sample, comprising the Dutch sample (53) and a random sample of 53 people from the UK sample was therefore deemed adequate for the purpose of this analysis.

Results

Indoor autonomy

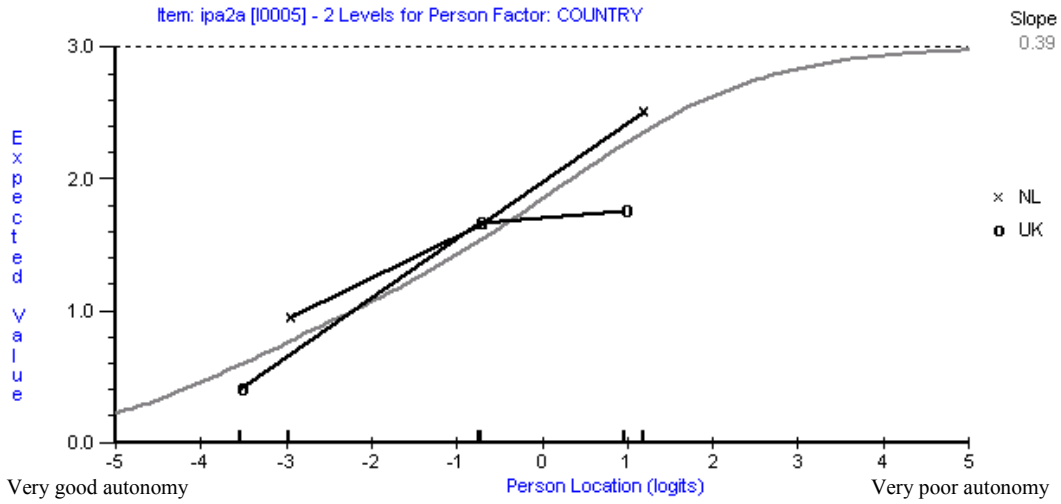
Three of the seven Indoor Autonomy items had disordered thresholds (items 2a, 2d and 2e). For each of these items the response categories 3 and 4 were combined, after which there was ordering of the thresholds. After rescoring, the data fitted the Rasch model, as indicated by the insignificant chi-square ($\chi^2 = 11.96$, P = 0.61). However, DIF analysis showed significant bias by country on item 2a (self-care achieved the way one chooses), with Dutch people reaching higher expected values than English people (figure 1 & table 1).

Table 1 Differential Item Functioning statistics after rescoring of items

Subscale items	**Uniform DIF gender p-value**	**Uniform DIF Country p-value**
Autonomy indoors		
1a	0.4203	0.3102
1b	0.2064	0.5760
2a	0.8986	0.0036
2b	0.6688	0.8066
2c	0.5364	0.7940
2d	0.1380	0.1086
2e	0.7759	0.7582
Family role		
3a	0.8931	0.0384
3b	0.2419	0.9204
3c	0.1487	0.9727
3d	0.0442	0.0348

3e	0.5648	0.9343
3f	0.1895	0.1524
4a	0.8007	0.0206
Autonomy outdoors		
1c	0.9551	0.1151
1d	0.2285	0.7010
5a	0.8109	0.3283
6g	0.8515	0.7338
10	0.2650	0.2068
Social Life and Relationships		
6a	0.3752	0.1693
6b	0.3466	0.7572
6c	0.7435	0.5703
6d	0.3637	0.2736
6e	0.8065	0.5922
6f	0.8352	0.0352
Work and Education		
7a	0.0321	0.6986
7b	0.3589	0.3466
7c	0.6023	0.7205
7d	0.3533	0.9885
7e	0.6329	0.9974
8a	0.0041	0.2287

Figure 1 Item Characteristic Curve for IPA Item 2a, displaying uniform DIF by country



Legend to figure 1: The figure displays an Item Characteristic Curve for item 2a. The x-axis shows the person locations, with lower scores reflecting better levels of autonomy and vice versa. The y-axis shows the expected response values; as the data was rescored this goes up to 3 and not 4. The grey line plots all the people in the sample, the blue line represents Dutch people, the red line English people: at any given level of autonomy Dutch people score higher (or poorer autonomy) than English people.

Item 2a was therefore split by country, resulting in two new items (2a NL & 2a UK). There was no significant non-uniform DIF on any of the indoor autonomy items. After splitting item 2a, the summary fit statistics indicated that the subscale fitted the Rasch model and the PSI was greater than 0.8. The item fit statistics were all within the acceptable range of -2.5 to 2.5 (table 2). Thus, the indoor autonomy subscale was

stable across the two cultures. The total variance accounted for, after removing the 'Rasch Factor' was 43.7%.

Table 2 Final Model fit statistics

Subscale	**Item fit statistics**		**Item fit residual**	**Person fit residual**	**Item trait interaction**		**Person Separation Index**
	Item fit residual	p-value	Mean (SD)	Mean (SD)	χ^2	p-value	
Autonomy indoors			0.098 (0.829)	-0.483 (1.366)	16.034	0.4506	0.9499
1a	0.352	0.6766					
1b	0.578	0.7467					
2b	-1.184	0.3650					
2c	-1.193	0.7810					
2d	0.152	0.7589					
2e	0.509	0.1579					
2aNL	0.584	0.2096					
2aUK	0.990	0.0992					
Family role			-0.013 (1.083)	-0.411 (1.189)	8.750	0.2711	0.9431
3a	-0.716	0.297264					
3b	0.341	0.956264					
3c	0.153	0.633686					
3d	-0.892	0.623451					
3e	-0.344	0.854385					

3f	-0.826	0.077752					
4a	2.189	0.044243					
Autonomy outdoors			0.439	-0.293	4.358	0.9298	0.912
			(0.952)	(1.041)			
1c	-0.915	0.3863					
1d	-0.205	0.8668					
5a	1.243	0.4822					
6g	1.133	0.7428					
10	0.940	0.9434					
Social Life and Relationships			-0.039	-0.420	4.254	0.6424	0.924
			(1.405)	(1.095)			
6a	-1.793	0.472420					
6b	-0.944	0.409714					
6c	0.019	0.774884					
6d	-0.595	0.208340					
6e	0.952	0.631713					
6f	2.125	0.280752					
Work and Education			0.150	0.376	2.262	0.8941	0.871
			(0.780)	(1.115)			
7a	-0.093	0.427326					
7b	0.631	0.672794					
7c	0.417	0.538390					
7d	-0.729	0.617762					
7e	0.659	0.726443					
8a	1.276	0.401529					

Family role autonomy

Two items of the 7-item family role subscale required rescoring (3d and 4a). There was no uniform or non-uniform DIF by gender or country (table 1). The summary fit statistics and item fit statistics showed that this subscale fitted the Rasch model (after rescoring items 3d and 4a) (table 2). The total variance accounted for, after removing the 'Rasch Factor' was 36.3%.

Outdoor autonomy

There were no disordered thresholds in this subscale. There was no uniform or non-uniform DIF by gender or country and there was no significant deviation from the Rasch model (tables 1&2). The PCA of the residuals showed that the total variance accounted for was 36.5%.

Social life and relationships

Two of the six items had disordered thresholds (items 6c and 6f). After rescoring these two items the data fitted the model and there was no uniform or non-uniform DIF by gender or country (table 1). The total variance accounted for was 40.8%.

Work and education

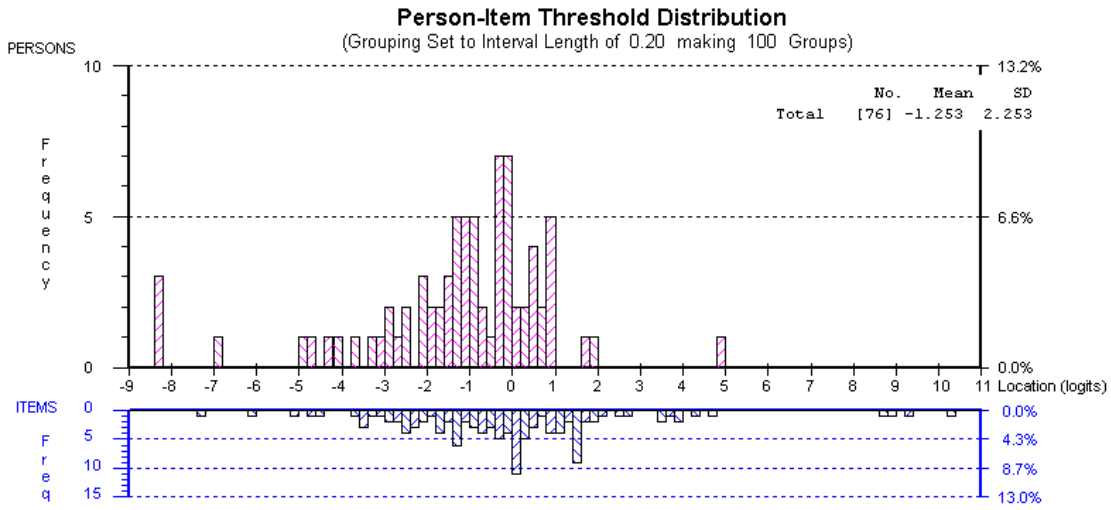
Three items required rescoring for this subscale. Although there were more missing data for this subscale as many people in the study did not work, there was no significant deviation from the Rasch model (table 2). Item 8a showed significant uniform and non-uniform DIF by gender but considering the small numbers we were unable to explore this further or adjust for it. The total variance accounted for was 35.1%.

31-item IPA

The total scale had 11 disordered thresholds which were all rescored successfully. One item displayed non-uniform DIF by country (item 6e; P<0.01) and was removed. One item showed significant DIF by country (item 4a; P<0.01) and this item was split by country. Similarly, one item showed DIF by gender (item 8a; P<0.01) and this item was split by gender. Following these procedures, the data fitted the Rasch model (χ^2 =82.13, P=0.06, Item Fit residual mean [SD] 0.22 [1.09]; Person fit residual mean [SD] 0.10 [1.63]; person separation index = 0.98). The principal component analysis was conducted on the dataset prior to splitting items 4a and 8a since the software does not allow this analysis to be conducted if items have been split. The total variance accounted for, after removing the 'Rasch Factor' was 14.6%. The two tests examining the assumption of local independence further were both insignificant, further supporting the scale's unidimensionality (t-test of positive factor loadings against the whole scale t=0.170; P=0.8652 and t-test of negative factor loadings against the whole scale t=0.538; P=0.592).

Figure 2 shows the item thresholds distribution (lower section of the figure) and the person distribution (upper section of the figure) of the 30-item scale. The graph shows that the item thresholds are well distributed along the new ruler of participation and autonomy (the x-axis). It can also be seen that there are few people with very poor participation and autonomy (located at the right end of the ruler).

Figure 2 Person-Item Threshold Map Distribution: 30-item IPA (item 6e removed)



Legend to figure 2: The x-axis shows the person locations (top half of the diagram) and item thresholds location (bottom half of the diagram), with lower scores reflecting better levels of autonomy and vice versa. The y-axis shows the number of people and item thresholds located at particular points of the scale.

Discussion

We tested the IPA in samples of people with varying levels of disability, drawn from populations based in the Netherlands and the United Kingdom. Rasch analysis showed that the IPA subscales and a 30-item IPA (as one item was removed) were invariant across the two cultures.

Sample characteristics

The data used in this analysis came from two separate observational studies, which employed different recruitment procedures. The samples were compatible in that the

age ranges were identical, participants were competent in the questionnaire's language and questionnaires were self-completed. Both included participants drawn from out-patient clinics. In addition, the English sample included patients recruited at GP practices. They were included in the cross-cultural validity analysis, in order to assess the stability of the instrument across disease index groups and also to evaluate discriminant validity, which was satisfactory.[9]

Formal DIF analysis by disease index group was not possible, however, due to the small numbers in the groups. This will be formally tested in further work.

Scale

It is recommended that the adaptation of scales is performed according to strict guidelines.[10] This ensures that the semantics of the questions remain the same after the questionnaire has been translated. Many scales are translated without proper evaluation, resulting in scales that are not valid for use across populations. Even those scales that are adapted properly, often do not formally test whether there is bias between countries, or, in other words, whether people from one country systematically answer a question differently from people in another country (i.e. whether there is bias by country). The English IPA language adaptation was a lengthy but thorough process, involving a number of bilingual researchers and lay people, experts and the original developer of the IPA. The Rasch DIF analysis adds to this by testing for invariance across cultural groups. Bias was found in one item only. This detailed exploration allowed the creation of two separate items from this question, thus locating people more accurately along the parametric ruler. This was a small

study, however, and bias by country needs to be examined in other, independent, samples.

The work and education subscale has not been formally validated in the Dutch studies as they included insufficient number of people who were in work. This analysis (and a previous confirmatory factor analysis[9]) has shown that this subscale was stable across country and largely gender groups, although one item (about educational opportunities) displayed uniform and non-uniform DIF by gender. This requires further investigation in studies with larger numbers of people that intend to follow education or already follow education.

Since the publication of the Dutch data[6], one question has been added to the IPA questionnaire. This question concerns people's participation and autonomy with respect to supporting and helping other people. This was deemed an important addition to the questionnaire which had not before included questions on reciprocity. This question has been added to the new English IPA and has been found to load onto the social life and relationships subscale in a confirmatory factor analysis.[9] However, we were unable to examine the cross-cultural validity of this question as it was not included in the Dutch data.

The final analysis, examining the validity of the 31-item IPA, showed that this scale fitted the Rasch model (after removing one item and adjusting for bias for two items). Further, the PCA results for the five subscales were all above the acceptable value of 30% whereas the PCA results for the entire 31-item scale were acceptable (<30%).[19] These results suggest that there is an overall higher level construct of participation and

autonomy and that data may not need to be analysed separately for each subscale. However, for clinical practice the subscale scores may be instrumental in identifying the focus for rehabilitation.

Analysis used

The IPA construct validity has previously been confirmed with a confirmatory factor analysis.[9] However, traditional psychometric methods do not examine invariance across groups (such as gender or country) or whether there is an ordered continuum of items that represent a unidimensional construct. Rasch analysis is therefore a preferred method for cross-cultural validity studies.

In this analysis we examined cross-cultural validity by pooling data and calibrating the IPA for both countries combined. Some calibrate data for countries separately (see for example the WHOQOL-100 study[23]). Had we used that method, we would have derived two separate calibrated rulers for participation and autonomy (i.e. a Dutch and English ruler). This would have enabled comparisons of item threshold locations, their difficulty levels and similarities between the two countries. However, it would not have been possible to make direct comparisons of the relative distance between the Dutch and English item thresholds, since the data would not have been calibrated onto a single ruler. DIF analysis is now becoming more standard practice for the use of cross-cultural validity studies and in our analysis it enabled us to make valid judgements about possible country biases.

Conclusions

Our analysis has shown that the IPA subscales and a 30-item IPA are invariant across two cultures (the Netherlands and the United Kingdom) and across gender. The 30-item IPA can therefore be used validly to assess participation and autonomy in these populations. Considering the good results for the 30-item IPA we suggest that when the IPA is used as an outcome measure it can be used as a whole, without the need to calculate subscale scores. However, when the IPA is used as an assessment tool the subscale scores may be instrumental in identifying the focus for rehabilitation.

Cross-cultural validity of the IPA needs to be examined each time another country translates the questionnaire for use locally. In addition, further analyses are required to examine whether the IPA is invariant across differing levels of disability and other disease groups not included in this study. However, our findings suggest that the IPA measures a construct that is likely to be found to be valid across populations sharing the general cultural characteristics of our Netherlands and UK samples.

Acknowledgements

The original work in the UK was funded by the NHS Executive South East R&D department. The original work in the Netherlands was performed at the time M. Cardol was employed in the Academic Medical Centre, the Netherlands. The authors are grateful to Mrs Ria Kersten for her assistance with the translation of the questionnaire and the participants for taking part in the studies.

References

1. Cardol M, de Jong BA, Ward CD. On autonomy and participation in rehabilitation. Disability & Rehabilitation 2002;24(18):970-4.

2. Cardol M, de Haan RJ, van den Bos GAM, de Jong BA, de Groot IJM. The development of a handicap assessment questionnaire: The Impact on Participation and Autonomy (IPA). Clin Rehabil 1999;13(5):411-9.

3. Cardol M, de Haan RJ, de Jong BA, van den Bos GAM, de Groot IJM. Psychometric properties of the impact on participation and autonomy questionnaire. Arch Phys Med Rehabil 2001;82(2):210-6.

4. Cardol M, Brandsma JW, de Groot IJ, van den Bos GA, de Haan RJ, de Jong BA. Handicap questionnaires: what do they assess? Disability & Rehabilitation 1999 Mar;21(3):97-105.

5. Cardol M, de Jong BA, van den Bos GAM, Beelen A, de Groot IJM, de Haan RJ. Beyond disability: Perceived participation in people with a chronic disabling condition. Clin Rehabil 2002;16(1):27-35.

6. Cardol M, Beelen A, van den Bos GA, de Jong BA, de Groot IJ, de Haan RJ. Responsiveness of the impact on participation and autonomy questionnaire. Arch Phys Med Rehabil 2002 Nov;83(11):1524-9.

7. Bosma H, Diederiks JP, van Santen HM, van Eijk JT. More social exclusion of chronically ill patients with lower incomes. [Dutch]. Nederlands Tijdschrift voor Geneeskunde 2005 Aug;149(34):1898-902.

8. Larsson Lund M, Nordlund A., Nygard J, Bernspang B. Perceptions of participation and predictors of perceived problems with participation in persons with spinal cord injury. J Rehab Med 2005;37(1):3-8.

9. 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; In press.

10. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 2000;25(24):3186-91.

11. Rasch G. Probablistic models for some intelligence and attainment tests. Chicago: University Chicago Press; 1980.

12. Andrich D. Rasch models for measurement. London: Sage; 1988.

13. Andrich D. Rating formulation for ordered response categories. Psychometrica 1978;47:149-74.

14. Fischer WP. Reliability statistics. Rasch Measurement Transactions 1992;6:238.

15. Scheuneman JD. A method of assessing bias in test items. J Educ Measure 1979;16:143-52.

16. Holland PW, Wainer H. Differential Item Functioning. Hillsdale. NJ: Lawrence Erlbaum; 1993.

17. Angoff WH. Perspectives on differential item functioning methodology. In: Holland PW, Wainer H, editors. Differential item functioning. Hillsdale. NJ: Lawrence Erlbaum; 1993.

18. Dorans NJ, Holland PW. DIF detection and description: Mantel-Haenszel and standardisation. In: Holland PW, Wainer H, editors. Differential item functioning. Hillsdale. NJ: Lawrence Erlbaum; 1993.

19. Smith EV. Detecting and Evaluating the Impact of Multidimensionality using Item Fit Statistics and Principal Component Analysis of Residuals. Journal of Applied Measurement 2002;3(2):205-31.

20. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995;310(6973):170.

21. RUMM 2020. Perth: RUMM Laboratory; 2003.

22. Linacre JM. Sample size and item calibration stability. Rasch Measurement Transactions 2006;7:28.

23. Leplege A, Ecosse E, WHOQOL Rasch Project Scientific Committee. Methodological issues in using the Rasch model to select cross culturally equivalent items in order to develop a Quality of Life index: the analysis of four WHOQOL-100 data sets (Argentina, France, Hong Kong, United Kingdom). Journal of Applied Measurement 2000;1(4):372-92.