

# Evaluation of the Social Participation Questionnaire in adult patients with depressive symptoms using Rasch analysis

Konstancja Densley · Sandra Davidson ·  
Jane M. Gunn

Accepted: 10 January 2013 / Published online: 23 January 2013  
© Springer Science+Business Media Dordrecht 2013

## Abstract

**Purpose** The aim of this study was to explore the psychometric properties of the 22-item Social Participation Questionnaire (SPQ).

**Methods** The SPQ was administered to 789 adult primary care patients with depressive symptoms. As the items were intended to be summed together to provide total score, Rasch analysis (partial credit model) was applied to assess the overall fit of the model, individual item fit, differential item functioning (DIF), targeting of persons, response dependency, unidimensionality and person separation.

**Results** To improve the scale's fit, it was necessary to re-score the response format. Two items demonstrated some DIF for gender and eight items showed DIF for age. To support the assumption of unidimensionality post hoc principal component analysis was performed. The analysis showed two subtests of the residuals with positive and negative loadings, but the person estimates derived from these two subtests were not statistically different to that derived from all items taken together. The response dependence between two items was identified; however, the magnitude of difficulty was very small. Although the questionnaire appeared to have insufficient items to assess the full spectrum of informal social contact, the SPQ was reasonably well targeted.

**Conclusion** The SPQ is a promising questionnaire for the measurement of social participation although it could benefit from the inclusion of further items to measure informal social contact. This study found support for the

internal validity, internal consistency reliability, and unidimensionality. A future study will investigate whether targeting can be improved when additional items are included.

**Keywords** Quality of life · Depression · Social participation · Rasch analysis · The Social Participation Questionnaire (SPQ)

## Introduction

Social participation is an indicator of an individual's integration in their community and is an important component of quality of life. Studies have shown that people who are well integrated in a group or community are buffered to some degree from the impact of daily stressors and non-routine crises that characterize human life [1–8]. This research suggests that social participation may have a legitimate place in health promotion and public policy, especially in relation to mental health [9]. However, to confidently establish the importance of social participation for health, and to identify which aspects of social participation are most critical, further research on the psychometric properties of existing social participation measures is necessary.

This study aims to examine the internal validity of a modified version of the Social Participation Index (SPI) [2] using Rasch analysis. The modified measure—the Social Participation Questionnaire (SPQ)—was tested to assess how well responses on the SPQ fit with the expectations of the measurement model. The analysis evaluates the validity of the category scoring system, the fit of individual items, and assesses of the potential bias of items by response dependence, gender and age.

---

K. Densley (✉) · S. Davidson · J. M. Gunn  
General Practice and Primary Health Care Academic Centre,  
The University of Melbourne, 200 Berkeley Street, Carlton,  
VIC 3053, Australia  
e-mail: kdensley@unimelb.edu.au

**Table 1** Final fit of the revised questionnaire items to the Rasch analysis

No.	Item	Loc.	SE	Fit res.	df	ChiSq.	df	Prob	Scoring					
1	Visited family/family visit <sup>b</sup>	-2.89	0.12	1.05	747.20	12.05	9	0.21	0	0	1	1	1	1
2	Visited friends/friends visit <sup>b</sup>	-2.82	0.11	-1.49	742.43	10.18	9	0.34	0	0	1	1	1	1
3	Visited neighbours/had neighbours visit <sup>b</sup>	-1.77	0.09	1.51	747.20	20.35	9	0.02	0	1	1	1	1	1
4	Attended church or a religious group <sup>c</sup>	0.16	0.08	0.92	748.15	5.24	9	0.81	0	1	1	1	1	1
5	Gone to a social club <sup>c</sup>	-0.15	0.08	-2.40	745.29	26.57	9	0.00	0	1	1	1	1	1
6	Used the internet for social communication <sup>ab</sup>	0.18	0.08	0.84	745.29	2.74	9	0.97	0	1	1	1	1	1
7	Gone to a café or restaurant <sup>c</sup>	-3.47	0.14	-1.64	747.20	9.91	9	0.36	0	1	1	1	1	1
8	Gone to a club, pub or bar <sup>ac</sup>	-1.80	0.09	-0.38	745.29	3.71	9	0.93	0	1	1	1	1	1
9	Gone to the cinema or theatre <sup>c</sup>	-2.06	0.09	-2.12	746.25	16.99	9	0.05	0	1	1	1	1	1
10	Gone to watch a sports event <sup>ac</sup>	0.30	0.08	-0.60	748.15	8.26	9	0.51	0	0	1	1	1	1
11	Gone to a party or dance <sup>c</sup>	0.07	0.08	-0.96	748.15	12.06	9	0.21	0	0	1	1	1	1
12	Played sport <sup>d</sup>	1.05	0.09	-1.34	749.11	10.49	9	0.31	0	0	1	1	1	1
13	Had social contact through children's sport <sup>ad</sup>	0.27	0.08	0.85	747.20	7.15	9	0.62	0	1	1	1	1	1
14	Gone to the gym or exercise class <sup>d</sup>	0.35	0.08	-0.81	745.29	7.80	9	0.55	0	1	1	1	1	1
15	Self-help or support groups <sup>d</sup>	1.19	0.10	-0.48	744.34	6.73	9	0.67	0	1	1	1	1	1
16	Singing, acting or played music in a group <sup>d</sup>	2.09	0.13	-0.88	750.06	4.20	9	0.90	0	1	1	1	1	1
17	Been involved in a hobby group <sup>d</sup>	0.88	0.09	-0.11	746.25	7.28	9	0.61	0	1	1	1	1	1
18	Gone to a class <sup>d</sup>	0.15	0.08	-0.27	747.20	12.85	9	0.17	0	1	1	1	1	1
19	School-related group <sup>c</sup>	1.50	0.11	-0.33	742.43	5.03	9	0.83	0	1				
20	Service club (e.g. Lions, CWA) <sup>c</sup>	2.62	0.16	-0.88	740.53	4.96	9	0.84	0	1				
21	Ethnic group (e.g. Croatian, Italian club) <sup>c</sup>	3.63	0.24	-0.41	735.76	6.65	9	0.67	0	1				
22	Volunteer organization or group <sup>c</sup>	0.50	0.09	0.41	747.20	8.42	9	0.49	0	1				

Items 1–18 were rescored from response category format: 012345 ('Once a week or more', 'A few times a month', 'Monthly', 'A few times a year', 'Rarely', 'Never'). Items 19–22 have dichotomous response format: 01 ('Yes', 'No')

<sup>a</sup> Items added

<sup>b</sup> Informal social contact

<sup>c</sup> Social contact through activities in public spaces

<sup>d</sup> Participation in group activities

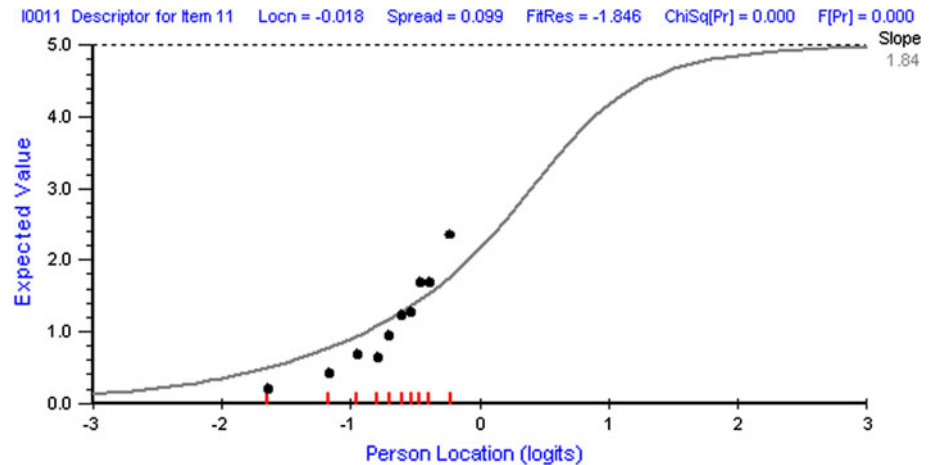
<sup>e</sup> Participation in community groups—mix of social and civic

## Method

The SPI was developed for use in The Adelaide Health Development and Social Capital Study [2]. Items in the SPI were selected by Baum et al. [2] on the basis of face validity and their knowledge of the social capital literature. They adopted a broad interpretation of social participation, and the SPI includes 29 items from the following six categories: informal social contact (three items; e.g. visiting family members); social contact through activities in public spaces (four items; e.g. a visit to a café or restaurant); participation in group activities such as sports or hobbies (six items; e.g. gym or exercise class); participation in civic activity undertaken on an individual basis (seven items; e.g. signing a petition or writing to a politician); civic participation involving group activity (four items; e.g. trade union or residents action group); and participation in community groups that involve a mix of social and civic purposes (five items; e.g. service clubs). The SPI has not undergone psychometric testing.

In the current study, the SPI was modified by removing two categories and adding four new items. Participation in civic activity undertaken on an individual basis and civic participation involving group activity were both removed from the questionnaire on the basis that earlier work of Baum and her colleagues found that endorsement for these categories was very low, and that they had a little association with mental health [2, 10]. Two items (Gone to a club, pub or bar and Gone to watch a sports event) were added to the section on social contact through activities in public spaces (Table 1), and one item was added to the section on participation in group activities (Had social contact through children's sport). Finally, one item was added to the category informal social contact asking about using the internet for social communication. These additional items were drawn from a list of social activities collated by Australian Bureau of Statistics and were considered to capture social activities frequently undertaken by Australian adults [11].

**Fig. 1** Item characteristic curve for item 11



The modified 22-item version of the questionnaire was called the SPQ (Table 1). Eighteen items in the SPQ asked respondents to indicate how regularly they had undertaken each activity in the last 12 months. They could choose one of the six response categories: “Never”, “Rarely”, “A few times a year”, “Monthly”, “A few times a month”, and “Once a week or more”. The remaining four items requested a binary “Yes”/“No” response regarding participation in community groups that involved a mix of social and civic purposes in the last 12 months.

Participants were 789 adults (226 males and 563 females) who (1) agreed to participate in an observational cohort study of depression in general practice [12] and (2) recorded a score of 16 or more on the Center for Epidemiologic Studies Depression Scale (CES-D) [13] in an initial screening survey. The CES-D is a widely used, short, self-report scale designed to measure depressive symptoms. The cut point of 16 provides the best discriminative ability to detect significant depressive symptoms in both community and patient samples [14]. At least 2 weeks after the screening survey, participants completed a self-administered questionnaire which included the SPQ. Because the SPQ items were intended to be summed together to provide total score, Rasch analysis (partial credit model) was applied to assess the psychometric properties of the scale [15].

#### Analytical methods

Data were fitted to the Rasch model using Rumm 2030 software [16]. Because the data adequately fit the Rasch model, simple comparisons of the items and respondents were possible. Analysing the differences of the estimates of the item parameters were assumed to be independent of the distribution of the participants being surveyed, however, the power of the test of fit may be affected by Rasch reliability measured through the person separation index [15].

The overall fit statistics were considered. The item–person and the item–trait interaction statistic were both transformed to approximate a z score, and it was reported as a Chi-square which reflected the property of invariance across the trait. Because the number of response categories varied across items, the item thresholds were not equal, and therefore, the partial credit parameterization was used to test the data [17].

## Results

### Overall fit

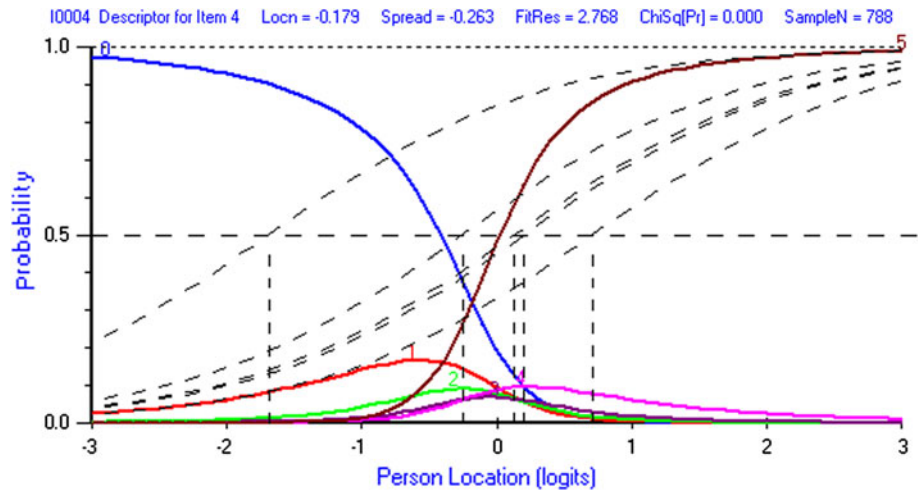
As the residual misfit for individual persons may skew the analysis due to their aberrant response patterns, misfitting persons were examined. There were only two respondents with a fit residual outside of the range  $\pm 2.5$ , and all results were confirmed by excluding them from the analysis.

Initial inspection of the scale showed poor overall fit to the Rasch model as evident in the standardized item fit residual statistic (mean =  $-0.555$ , SD =  $1.857$ ) and the item–trait interaction statistic ( $\chi^2 = 473.447$ ,  $df = 198$ ,  $p < 0.001$ ). This suggests that the relative operation of the items was not consistent across all levels of the underlying trait and that there was misfit between the data and the model [18]. The overall fit of the model was confirmed by inspection of the individual item fit statistics revealing misfitting items and graphs of the Item Characteristics Curve (Fig. 1).

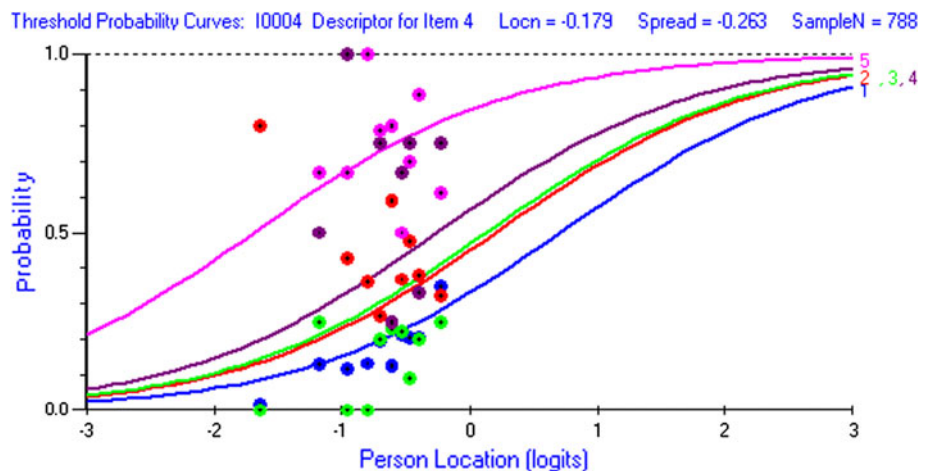
### Thresholds

The pattern of thresholds was examined to assess whether disordering might affect the fit. Initial inspection showed that the thresholds of most items were disordered. Exceptions included the four dichotomous items and items 7 and 8. Disordered thresholds occur when respondents have difficulty

**Fig. 2** Category probability curve for item 4 before rescoring showing disordered thresholds



**Fig. 3** Threshold probability curve for item 4 before rescoring where thresholds did not discriminate



consistently discriminating between response options. This can occur when there are too many response options, or when the labelling of options is potentially confusing or open to misinterpretation [19].

For responses to items with disordered thresholds, the category and thresholds probability curves were inspected (Figs. 2, 3). For a well-fitting item, it would be expected that across the entire range of the trait, each response category would systematically take turns showing the highest probability of endorsement. Inspection of the category and thresholds probability curves for all 22 items confirmed that none of the thresholds operated as required [18]. The thresholds and category probability curves for items 10 (Gone to watch a sports event), 11 (Gone to a party or dance) and 12 (Played sport) suggested that respondents were not able to reliably distinguish the response categories “Never” and “Rarely”.

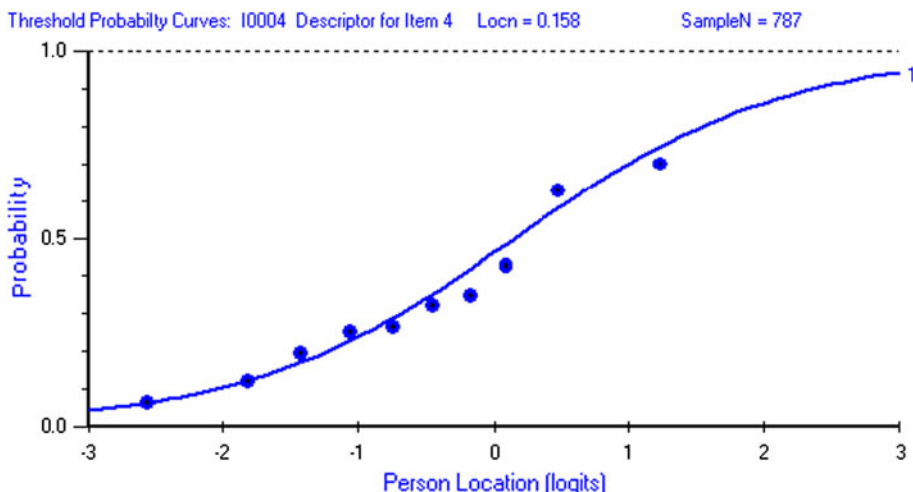
Disordered thresholds can be remedied by combining adjacent categories and retesting the fit of the data to the Rasch model. Although excessive collapsing is not recommended

[20], it can be justified in cases if discrimination at a pair of thresholds is close to zero and threshold estimates are reversed from their natural order [21].

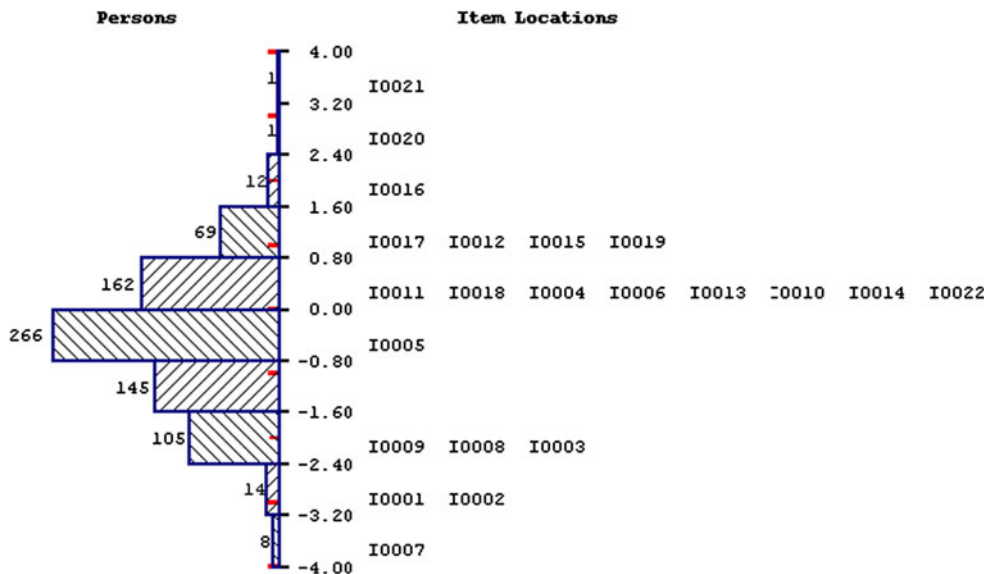
Items 1–18 were rescored one at a time (Table 1). After a series of separate analyses, this procedure resulted in an improvement of the overall model fit, with a change in the item–trait interaction probability value from 0.000 to 0.272. Although this represents a change to the original format of the questionnaire, it more closely represents the actual response patterns of the respondents in this sample. It was confirmed by inspecting threshold probability curves (Fig. 4) which showed that rescoring resulted in improvement in the functioning of the item with well-discriminating threshold.

The rescoring of these items suggests that the questionnaire may potentially be modified by reducing all responses categories to a binary format. It is possible that “Yes”/“No” response categories for the question: ‘In the past 12 months, have you done any of the following activities’ might be more valid than using a multiple

**Fig. 4** Threshold probability curve for item 4 after rescored where thresholds operated successfully



**Fig. 5** Person item map showing distribution of persons for revised SPQ



response format. Exceptions apply to 5 items: Visited family/family visit; Visited friends/friends visit; Gone to watch a sports event; Gone to a party or dance and Played sport, where more specific response categories such as “Monthly or more”/“Less than monthly” would be more appropriate.

**Overall fit after items rescored**

Rescoring improved the overall model fit, with a change in the residual standard deviation value for items from 1.857 to 1.043 and a change in the item–trait interaction total Chi-square probability value from 0.00 to 0.27. Relative operation of the items was consistent across all levels of the underlying trait and overall the data fit the model [18]. An estimate of the internal consistency reliability of the scale

based on the person separation index ( $PSI = 0.72$ ) indicated that the questionnaire had reasonable person separation reliability. Therefore, it had reasonable power to detect misfit and indicated that the items measured an underlying (or latent) construct and is appropriate for group use [22].

The residual mean value for persons was  $-0.282$  with a SD of 0.739, indicating no misfit among the respondents in the sample and higher discrimination than expected by the model. The residual mean value for items was  $-0.433$  with a SD of 1.043, close to the expected value of 1, suggesting no misfitting items. Alternative recoding procedures were also checked, however, no other solution improved the value of Fit Residual Standard Deviation.

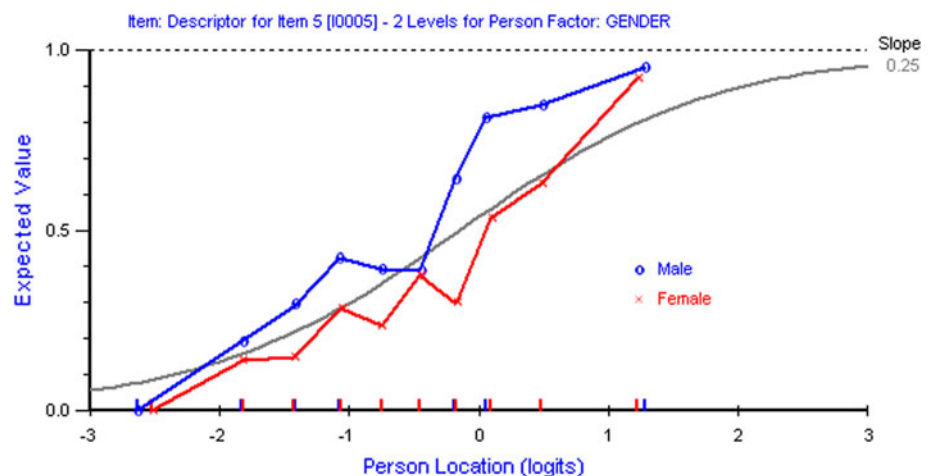
Finally, for the purpose of cross-validation, the model was confirmed using data provided by the same people who

**Table 2** Uniform and non-uniform DIF statistics for gender

Item	Uniform DIF				Non-uniform DIF			
	MS	F	df	Prob	MS	F	df	Prob
1	1.12	1.00	1	0.32	1.18	1.05	9	0.40
2	1.20	1.59	1	0.21	1.30	1.73	9	0.08
3	4.63	4.36	1	0.04	1.04	0.98	9	0.46
4	0.15	0.15	1	0.70	1.00	0.98	9	0.45
5	12.22	15.35	1	0.00	1.09	1.37	9	0.20
6	2.92	2.88	1	0.09	1.04	1.03	9	0.41
7	0.59	0.89	1	0.35	0.58	0.86	9	0.56
8	13.02	14.18	1	0.00	0.60	0.66	9	0.75
9	6.77	9.01	1	0.00	1.23	1.64	9	0.10
10	0.31	0.34	1	0.56	0.49	0.53	9	0.85
11	2.37	2.65	1	0.10	0.77	0.86	9	0.56
12	2.68	3.25	1	0.07	0.90	1.10	9	0.36
13	0.04	0.04	1	0.84	0.69	0.68	9	0.73
14	7.69	8.62	1	0.00	0.93	1.04	9	0.41
15	1.65	1.82	1	0.18	0.75	0.83	9	0.59
16	0.15	0.19	1	0.67	1.93	2.40	9	0.01
17	1.44	1.53	1	0.22	1.68	1.79	9	0.07
18	3.71	4.01	1	0.05	1.16	1.25	9	0.26
19	8.11	8.91	1	0.00	0.56	0.61	9	0.79
20	2.81	3.68	1	0.06	0.95	1.24	9	0.27
21	0.96	1.18	1	0.28	0.34	0.42	9	0.93
22	2.02	2.05	1	0.15	0.95	0.96	9	0.47

completed the SPQ 12 months later. This sample comprised 542 respondents (155 males and 387 females). The item–trait interaction statistic was  $\chi^2 = 207.070$ ,  $df = 176$ ,  $p = 0.05$ , and the residual mean value for items was  $-0.378$  with a standard deviation of 1.202. The fit residual mean value for persons was  $-0.287$  with a standard deviation of 0.692; the PSI was 0.71.

**Fig. 6** Differential item functioning graph of females and males for item 5



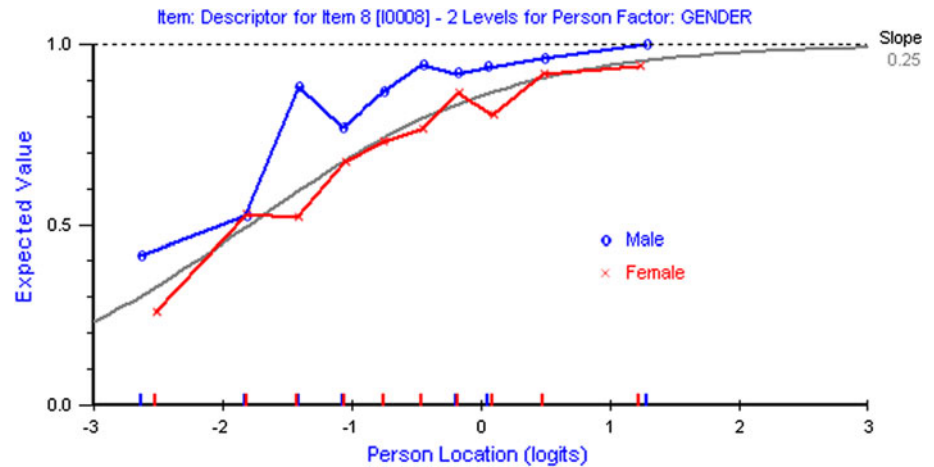
Targeting for Social Participation Questionnaire

It is important, particularly in clinical practice, that measures are appropriately targeted to the population being assessed. Poorly targeted measures may result in floor or ceiling effects. Figure 5 shows the frequency distribution of responses along the Rasch calibrated metric scale of social participation which provides an indication of how well targeted the items are for the study population. The left panel displays the frequency distribution of respondents along the Rasch calibrated metric scale of the attribute being measured (social participation). Cases at the bottom of the display have the lowest location value, representing low levels of social participation, while cases at the top have a high magnitude of social participation.

The right side of the graph displays the positions of each of the items. Items positioned at the bottom of the display are the easiest items to endorse. In this case, the most commonly endorsed item was item 7 (Gone to a café or restaurant) while the least commonly endorsed was item 21 (Ethnic group). The lower endorsement of the Ethnic group item can be seen in the light of the overall sample characteristics in which only 17 % were born in a country other than Australia. While attendance at an ethnic club is not restricted to people born outside Australia, it is specifically targeted to this group. In contrast, going to a café or restaurant is an activity that is equally (with the possible exception of financial disadvantage) available to all participants.

The scale shows a small gap in the graph between item 5 (Gone to a social club) and items 3 (Visited neighbours/had neighbours visit), 8 (Gone to a club, pub or bar), and 9 (Gone to the cinema or theatre) indicating that the questionnaire could benefit from inclusion of further items to measure lower level (informal) social participation of the group as a whole.

**Fig. 7** Differential item functioning graph of females and males for item 8



Overall, the average mean person location value of  $-0.548$  suggested that the questionnaire was reasonably well targeted (not too easy, not too hard) for use with this group. However, it was noted that some items were slightly less likely to be endorsed.

#### Individual item fit after items rescaling

Following the recoding of items, the overall fit of the model was confirmed by inspection of the individual item fit statistics (a summation of individual item deviations) revealed no misfitting items (Table 1). All items showed fit residual values within the range of  $-2.5$  to  $+2.5$ , with probability value of Chi-square more than 0.01, suggesting that the data fit the model. Item 5, Gone to a social club, showed probability value of Chi-square of 0.002. However, after adjusting the alpha value for the Bonferroni corrections (0.01), the Chi-square fit statistics for this item no longer indicated significant deviation from the model. A Bonferroni adjustment is typically applied to the alpha value used to assess statistical significance by dividing the alpha value of 0.01 by the number of items in the scale, in this case, adjusted alpha value was  $0.01/22 = 0.000455$  [23].

#### Differential item functioning

Having reached a satisfactory fit to the Rasch model with the rescored items of the SPQ, assessment of differential item functioning (DIF) was conducted using both statistical and graphical procedures. Results from a previous study [2] indicated that demographic variables such as gender and age were significantly associated with social participation levels, and therefore, DIF analysis was conducted to compare item location with respect to gender (male/female) and age (18–34/35–54/55–76).

According to the DIF analysis, two items showed probability values exceeding the adjusted alpha value (the Bonferroni-adjusted  $p$  value was 0.05) (Table 2). As there was no non-uniformity for all items (i.e. no interaction between the class intervals and gender), uniform DIF for item 5 and 8 (Figs. 6, 7) was interpreted directly [24]. Females were less likely than males to endorse these items (items 5 and 8). Because the difference was significant ( $p < 0.001$ ), these items were calibrated separately for males and females, treating it as two separate scale items [25].

As sequentially resolving items helps enforce the other items to show artificial DIF, this procedure was conducted for gender [25]. Because items 5 and 8 had a similar mean square, each one was resolved first in turn. Because item 8 had a higher mean square, it was split first. Following resolving item 8, DIF analysis using Bonferroni-adjusted  $p$  value of 0.05 for item 5 still showed significant uniform DIF ( $F(1, 788) = 13.91, p < 0.0001$ ).

Similarly, when item 5 was resolved first in turn, DIF analysis using Bonferroni-adjusted  $p$  value of 0.05 for item 8 still showed significant uniform DIF ( $F(1, 788) = 14.47, p < 0.0001$ ). After splitting items 5 and 8 together, all items did not show uniform or non-uniform DIF (using both Bonferroni-adjusted  $p$  values of 0.01 and 0.05). The logit difference for males and females in the split item 5 was 0.7 and in split item 8 it was 0.9. However, splitting of items 5 and 8 did not result in an improvement in the overall model fit. Overall, females were less likely to participate in social activities such as going to a club, pub, bar or social club. This is consistent with a previous study showing more social contact through activities in public spaces for males [2].

In terms of the age differences (Table 3), older participants were less likely to endorse SPQ items compared to those in younger age groups (Fig. 8). As there was no non-uniformity for all items (no interaction between the class intervals and age), uniform DIF for these items was interpreted directly.

**Table 3** Uniform and non-uniform DIF statistics for age

Item	Uniform DIF				Non-uniform DIF			
	MS	F	df	Prob	MS	F	df	Prob
1	2.33	2.08	2	0.13	0.84	0.75	18	0.76
2	0.42	0.55	2	0.58	0.77	1.01	18	0.44
3	18.35	18.46	2	0.00	2.08	2.09	18	0.01
4	6.18	6.12	2	0.00	0.62	0.61	18	0.89
5	10.02	12.77	2	0.00	1.03	1.31	18	0.17
6	7.67	7.77	2	0.00	1.38	1.40	18	0.12
7	0.17	0.26	2	0.77	0.64	0.95	18	0.52
8	1.65	1.78	2	0.17	1.11	1.21	18	0.25
9	0.62	0.82	2	0.44	1.11	1.47	18	0.09
10	10.46	11.72	2	0.00	0.83	0.93	18	0.54
11	20.65	24.27	2	0.00	0.57	0.68	18	0.84
12	2.75	3.36	2	0.04	0.93	1.13	18	0.31
13	23.06	24.00	2	0.00	0.71	0.74	18	0.77
14	5.52	6.15	2	0.00	0.54	0.60	18	0.90
15	1.29	1.43	2	0.24	1.25	1.39	18	0.13
16	4.86	6.07	2	0.00	0.97	1.22	18	0.24
17	4.23	4.44	2	0.01	0.21	0.22	18	1.00
18	2.58	2.80	2	0.06	1.30	1.41	18	0.12
19	5.58	6.23	2	0.00	1.24	1.39	18	0.13
20	11.00	14.81	2	0.00	0.70	0.95	18	0.52
21	0.83	1.03	2	0.36	0.66	0.82	18	0.68
22	12.68	13.27	2	0.00	0.94	0.98	18	0.48

To detect potential artificial DIF, a procedure of sequentially resolving items was conducted [25]. As item 11 had the highest mean square, it was split first and the overall model fit and individual item statistics was checked. Following resolving item 11, all items did not show uniform or non-uniform DIF (using both Bonferroni-adjusted  $p$  values of 0.01 and 0.05). The logit difference for resolved item 11 between participants in the oldest age group (55–75 years) and participants in the youngest age group (18–34 years) was almost one logit (0.9). The split items procedure resulted in a slight improvement in the overall model fit with PSI increased from 0.72 to 0.74. Overall, older participants were less likely to participate in social activities such as attending a party or a dance, which is consistent with previous research showing that social contact through activities in public spaces decreased as age increased [2].

### Unidimensionality

A principal component analysis was conducted on the residuals to identify the two most divergent subsets of items [26]. Items may load positively or negatively upon the first principal component and those items with the

highest loading in either direction may represent sets of items that are the most likely to breach the assumption of local independence. Examination of the item residuals matrix revealed no correlations greater than  $\pm 0.3$ , and the eigenvalue of 1.932 for the first component was only slightly larger than the eigenvalues for the other components [27]. The initial principal component analysis explained only 8.787 % of the total variance among residuals which suggested a small possibility that a second component existed [28].

Further analysis of the pattern of residuals showed that the residuals loaded in opposite directions on the first principal component (Table 4). This showed that these items have a common feature in addition to the factor structure that is common across all the items [26]. A clear distinction between the two subsets of the component was that items in the positively loaded set represented informal social activities while items in the negatively loaded set represented, with the exception of the visiting one's neighbour's, formal group participation. Participation in the informal activities may or may not involve other people, whereas the activities in the latter set are designed around a group format.

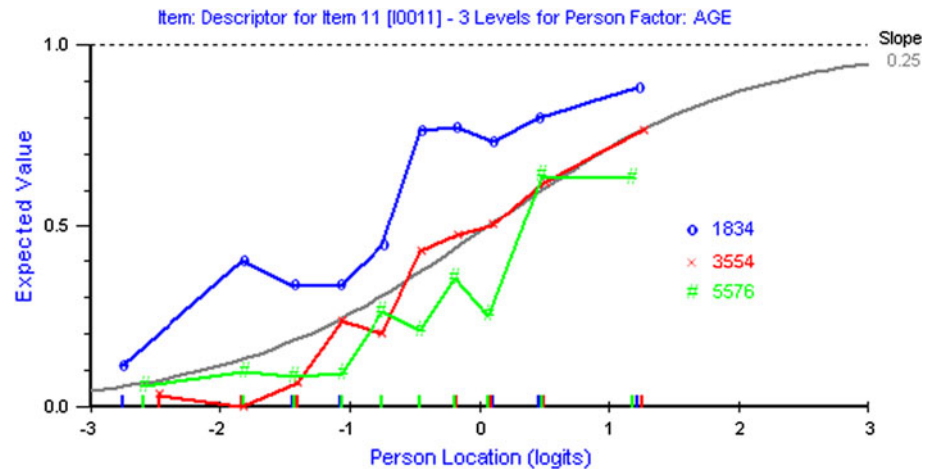
These two subsets of items were then separately fitted to the Rasch model and the person estimates obtained. The differences in person estimates derived from these analyses were trivial. The difference between the first subset and the overall scale was 0.07 logit, and the difference between the second subset and the overall scale was 0.03 logit. The reliability for the SPQ obtained from the unidimensional approach was much higher than reliabilities calculated from the two-dimensional approach.

A paired samples  $t$  test (at 5 % level of significance) was conducted to test whether person estimates derived from the two subsets of items were significantly different from each other. Corrected for the errors of measurement correlation between the two subsets was of the order of 0.36 and 0.35, and there was no statistically significant difference in the person estimates derived from the two subsets (set1:  $M = -0.557$ ,  $SD = 1.297$ ; set2:  $M = -0.504$ ,  $SD = 1.256$ ,  $M_{diff} = -0.053$ ,  $SD = 1.449$ ;  $t(789) = -1.027$ ). These results indicate that assumption of unidimensionality was not breached [22].

The person ability estimates based on the subsets of items were then generated and compared using a series of independent  $t$  tests (one for each case). Although the person locations estimated from these two subsets differed significantly for 6.5 % of participants at 5 % level of significance, the 95 % confidence intervals around this estimate, derived from a binomial distribution, included five per cent (CI: 4.8–8.4) which indicated that the unidimensionality of the 22-item SPQ was supported [26].



**Fig. 8** Differential item functioning graph for different age groups for item 11



**Table 4** Principal component analysis of the residuals showing loadings on the first component extracted

Item	Item	Loading
10	Gone to watch a sports event*	0.511
11	Gone to a party or dance	0.424
8	Gone to a club, pub or bar*	0.402
13	Had social contact through children's sport*	0.328
12	Played sport	0.245
1	Visited family/family visit	0.236
7	Gone to a café or restaurant	0.198
9	Gone to the cinema or theatre	0.163
14	Gone to the gym or exercise class	0.103
2	Visited friends/friends visit	0.087
19	School-related group	0.073
6	Used the internet for social communication*	0.054
3	Visited neighbours/had neighbours visit	-0.142
5	Gone to a social club	-0.161
15	Self-help or support groups	-0.225
21	Ethnic group (e.g. Croatian, Italian club)	-0.228
16	Singing, acting or played music in a group	-0.257
18	Gone to a class	-0.309
4	Attended church or a religious group	-0.385
20	Service club (e.g. Lions, CWA)	-0.412
17	Been involved in a hobby group	-0.434
22	Volunteer organization or group	-0.444

### Response dependence

The matrix of correlations from the standardized residuals showed that there was only one correlation that exceeded a value of 0.2 between items. These items were 10 (Gone to watch sports event) and 13 (Had social contact through children's sport). This correlation might indicate some form of response dependence.

Item 13 was closer to the mean of the person distribution ( $-0.548$ ) while Item 10 was slightly more difficult (Table 1). Therefore, item 10 was resolved, based on the response on item 13, and the value of estimated magnitude of response dependence between items 10 and 13  $d$  was calculated [29]. The value of estimated magnitude of response dependence between items 10 and 13  $d$  was highly significant at a conservative 1 % ( $>2.58$ ) level of significance ( $z = 0.613/0.045 = 13.699$ ). However, the magnitude of the difficulty caused by dependence was of the order of only 0.6 logits. Those participants who scored 0 ("Never") on item 13 (Had social contact through children's sport) found item 10 (Gone to watch sports event) more difficult to endorse, however, it was only 0.6 logits greater than if there was no dependence. The result of the analysis was confirmed using complete data set.

### Conclusion

The current study undertook a rigorous examination of the SPQ. Using an Australian sample of primary care patients with depressive symptoms, it would appear that the SPQ is a promising measure of social participation. However, the questionnaire could benefit from the inclusion of further items to measure lower levels of social participation. Rasch analysis revealed several valuable insights into the psychometric properties of the SPQ. Social Participation Questionnaire shows promising internal consistency and reliability which supports its use in studies including people with depressive symptoms.

Overall, the questionnaire was reasonably well targeted for use among adults with depressive symptoms, although the sample as a whole found some items slightly harder to endorse than others. Rasch analysis showed a single, uni-dimensional structure of the questionnaire. A rigorous test of the local independence assumption demonstrated a very

small magnitude of difficulty caused by dependence. Using the 12-month follow-up, sample for cross-validation of the Rasch-derived SPQ supported its utility and structural validity.

Further validation studies of the revised SPQ in the same cohort of adult patients with depressive symptoms will be conducted in the future to see whether additional items make it more appropriate for use with this group. These modifications should target the spectrum of informal social contact and include items such as: Gone for a walk with friends/family, Gone shopping with friends/family, Gone to museum or gallery with friends/family, or Talked to friends or family on the phone. In addition, because the initial scoring was not consistent with the model, modification of the questionnaire format could be trialled by simplifying all response categories to a binary format. More straightforward responses labels such as: “Yes”/“No” or “Monthly or more”/“Less than monthly” could be examined. It is possible that “Yes”/“No” response categories for the question: ‘In the past 12 months, have you done any of the following activities’ might be more valid than using a multiple response format. Exceptions apply to 5 items: Visited family/family visit; Visited friends/friends visit; Gone to watch a sports event; Gone to a party or dance and Played sport, where more specific response categories such as “Monthly or more”/“Less than monthly” would be more appropriate. Reducing the number of categories and simplifying their response labels will also decrease the burden on respondents and make interpretation of results easier. As the possibility of gender and age influencing responses to some of the items was demonstrated, the SPQ will also be tested to assess category ordering and the presence of DIF particularly in relation to age differences across clinical groups.

Facilitating social activities among adults is a promising direction for programs intended to promote mental health. Further validation of the questionnaire with additional items using the same sample will assess in depth the pattern of social participation in patients with depressive symptoms and might help in defining determinants of social (limited social activity) isolation.

**Acknowledgments** The named authors submit this publication on behalf of the diamond study investigators which include: Professor Jane Gunn, Professor Helen Herrman, Professor Mike Kyrios, A/Professor Kelsey Hegarty, Professor Christopher Dowrick, Dr. Gail Gilchrist, Associate Professor Grant Blashki, Professor Dimity Pond, Ms. Patty Chondros, Dr. Renata Kokanovic and Dr. Victoria Palmer. The diamond study was initiated with pilot funding from the beyondblue Victorian Centre of Excellence, and the main cohort has received project grant funding from the National Health and Medical Research Council (ID 299869, 454463 and 566511). The 1-year Computer-Assisted Telephone Interview was funded by a Stream 3 grant from the Australian Primary Health Care Research Institute (APHCRI). No funding body had a role in study design; the

collection, analysis, and interpretation of data; the writing of the manuscript; or the decision to submit this manuscript for publication. We acknowledge the 30 dedicated GPs, their patients and practice staff for making this research possible. We thank the cohort participants for their ongoing involvement in the study. We also thank the diamond project team and associate investigators involved in the study: Dr. Lena Sanci, Ms. Catherine Mihalopoulos, Ms. Maria Potiriadis, Ms. Aves Middleton and the casual research staff.

## References

1. Durkheim, E. (1951). *Suicide: A study in sociology* (J. A. Spaulding & G. Simpson, Trans.). New York: The Free Press.
2. Baum, F. E., Bush, R. A., Modra, C. C., Murray, C. J., Cox, E. M., Alexander, K. M., et al. (2000). The epidemiology of participation: An Australian community study. *Journal of Epidemiology and Community Health*, *54*(6), 414–423.
3. Berry, H. L., Rodgers, B., & Dear, K. B. G. (2007). Preliminary development and validation of an Australian community participation questionnaire: Types of participation and associations with distress in a coastal community. *Social Science and Medicine*, *64*(8), 1719–1737.
4. Kawachi, I., & Berkman, L. F. (2001). Social ties and mental health. *Journal of Urban Health*, *78*, 458–467.
5. Murray, G., Judd, F., Jackson, H., Fraser, C., Komiti, A., Pattison, P., et al. (2007). Ceremonies of the whole: Does social participation moderate the mood consequences of neuroticism? *Social Psychiatry and Psychiatric Epidemiology*, *42*, 173–180.
6. Phongsavan, P. F., Chey, T., Bauman, A., Brooks, R., & Silove, D. (2006). Social capital, socio-economic status and psychological distress among Australian adults. *Social Science and Medicine*, *63*, 2546–2561.
7. Isaac, V., Stewart, R., Artero, S., Ancelin, M. L., & Ritchie, K. (2009). Social activity and improvement in depressive symptoms in older people: A prospective community cohort study. *American Journal of Geriatric Psychiatry*, *17*(8), 688–696.
8. Chiao, C., Weng, L. J., & Botticello, A. (2011). Social participation reduces depressive symptoms among older adults: An 18-year longitudinal analysis in Taiwan. *BMC Public Health*, *11*(1), 292.
9. Israel, B., Checkoway, B., Schulz, A., & Zimmerman, M. (1994). Health education and community empowerment: Conceptualizing and measuring perceptions of individual, organizational and community control. *Health Education Quarterly*, *21*(2), 149–170.
10. Ziersch, A. M., & Baum, F. E. (2004). Involvement in civil society groups: Is it good for your health?. *Journal of Epidemiology and Community Health*, *58*, 493–500.
11. Australian Bureau of Statistics. (2004). *Measuring Social capital. An Australian Framework and indicator*, Cat. No. 1378.0. Canberra: Australian Bureau of statistics.
12. Gunn, J. M., Gilchrist, G. P., Chondros, P., Ramp, M., Hegarty, K. L., Blashki, G. A., et al. (2008). Who is identified when screening for depression is undertaken in general practice? Baseline findings from the diagnosis, management and outcomes of depression in primary care (diamond) longitudinal study. *The Medical Journal of Australia*, *188*, S119–S125.
13. Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401.
14. Thomas, J. L., Jones, G. N., Scarinci, I. C., Mehan, D. J., & Brantley, P. J. (2001). The utility of the CES-D as a depression screening measure among low-income women attending primary care clinics. *The International Journal of Psychiatry in Medicine*, *31*(1), 25–40.

15. Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum.
16. Andrich, D., Sheridan, B. E., & Luo, G. (2009). *RUMM 2030: Rasch multidimensional models for measurement*. Perth, WA: RUMM Laboratory.
17. Andrich, D. (2010). Understanding the response structure and process in the polytomous Rasch model. In M. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models: Developments and applications*, Chapter 6. New York: Routledge. Taylor & Francis.
18. Andrich, D. (2010). Educational measurement: Rasch models. In E. Baker, P. Peterson, & B. McGaw (Eds.), *International encyclopedia of education* (3rd ed.). Oxford: Elsevier.
19. Andrich, D. (1978). Rating formulation for ordered response categories. *Psychometrika*, *43*, 561–573.
20. Andrich, D. (2006). Item discrimination and Rasch-Andrich thresholds revisited. *Rasch Measurement Transactions*, *20*, 1055–1057.
21. Adams, R. J., Wu, M. L., & Wilson, M. (2012). The Rasch rating model and the disordered threshold controversy. *Educational and Psychological Measurement*, *72*(4), 547–573.
22. Tennant, A., & Conaghan, G. (2007). The Rasch measurement model in rheumatology: What is it? Why use it? When should it be applied and what should one look for in a Rasch paper? *Arthritis and Rheumatism*, *57*(8), 1358–1362.
23. Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *British Medical Journal*, *310*, p170.
24. Broderson, J., Meads, D., Kreiner, S., Thorsen, H., Doward, L., & McKenna, S. (2007). Methodological aspects of differential item functioning in the Rasch model. *Journal of Medical Economics*, *10*, 309–324.
25. Andrich, D. A., & Hagquist, K. (2011). Real and artificial differential item functioning. *Journal of Educational and Behavioural Statistics*, *37*(9), 387–416.
26. Smith, E. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, *3*, 205–231.
27. O'Connor, B. P. (2000). SPS, SAS, and MATLAB programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, and Computers*, *32*, 396–402.
28. Pallant, J. F., & Tennant, A. (2006). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology*, *46*, 1–18.
29. Andrich, D., & Kreiner, S. (2009). Quantifying response dependence between two dichotomous items using the Rasch model. *Applied Psychological Measurement*, *34*(3), 181–192.