Psychometric properties of the PsychoSomatic Problems scale – an examination using the Rasch model

Curt Hagquist

Karlstad University, Karlstad, Sweden

Address:
Karlstad University
SE-651 88 Karlstad
Sweden

Phone: + 46 54 7002536
Fax: + 46 54 7002523
Email: curt.hagquist@kau.se

Paper submitted to Objective Measurement in the Social Sciences

The ACSPRI Social Science Methodology Conference 2006
The University of Sydney, Sydney, Australia
10-13 December 2006

Psychometric properties of the PsychoSomatic Problems scale – an examination using the Rasch model

Abstract

The PsychoSomatic Problems (PSP)-scale is built upon eight items intended to tap information about psychosomatic problems among schoolchildren and adolescents in general populations.

The purpose of the study is to analyse the psychometric properties of the PSP-scale by means of the Rasch model, with a focus on the operating characteristics of the items. Cross-sectional data collected in Sweden at six points of time among adolescents is used for the analysis. In all more than 15 000 students are included in the analysis. Data were examined with respect to invariance across the latent trait, Differential Item Functioning (DIF) and item categorisation.

The results show that the PSP-scale adequately meets measurement criteria of invariance and proper empirical ordering of the data. Also the targeting is good and the reliability is high. Taking DIF into account by splitting problematic items with respect to gender provides a scale that shows no or only small signs of DIF. It is concluded that the PSP-scale is appropriate for measurement of psychosomatic health in general populations of adolescents.

Introduction

Child and adolescent mental health is an important issue on the public health agenda in many Western countries. Although deteriorations of mental health among young people are frequently reported, there are significant gaps in the knowledge about time trends and distributions of mental health problems across different sociodemographic groups of young people. Measurement problems also make comparisons between and within countries difficult. In psychiatric orientated epidemiological research among children and adolescents the Child Behavior Checklist (Achenbach, 1991a; 1991b) and the Strengths and Difficulties Questionnaire (Goodman, 2001) are the predominant instruments. In research within public health, the focus is rather on psychosomatic health than on psychopathology. There are various instruments on psychosomatic health available which share a lot in common. In Europe the HBSC-instrument is one of the most well known instruments in the field of psychosomatic health. The HBSC-instrument comprises eight questions and is considered to measure somatic as well as psychological problems. The reports of the results are also building on that distinction. A few studies have examined the psychometric properties of the instrument and there are uncertainties about the dimensionality as well as about other psychometric aspects of that instrument. A major problem recognised is how items are rated and categorised which is characterised by a mixture of quantitative and qualitative response categories (Hagquist & Andrich, 2004a). In contrast to the HBSC-instrument the PsychoSomatic Problems scale (the PSP-scale) has a response format based only on qualitative response categories. Similar to the HBSC-instrument the PSP-scale is built upon eight items intended to tap information on psychosomatic problems among children and adolescents in general populations. The

PSP-scale has so far mainly been used in local and regional studies among adolescents across Sweden.

The purpose of the study is to examine the psychometric properties of the PSP-scale by means of the Rasch model, with a focus on the operating characteristics of the items.

Methods

Material

The analysis is based on cross-sectional data collected at six points of time 1988-2005 among adolescents (15-16 years old) in year 9 within a county in Sweden.
This study makes use of data from those municipalities which have been participating all years of investigations (14 out of 16 municipalities). In all the data set comprises 15135 students. The number of respondents each year was as follows: 2701 (1988), 2605 (1991), 2426 (1995), 2342 (1998), 2455 (2002) and 2643 (2005). The corresponding attrition rates were: 10.0 % (1988), 10.9 % (1991), 6.3 % (1995), 9.0 % (1998), 11.8 (2002) and 14.3 (2005).

Data collection

The data were collected with a questionnaire, which was handed out in the classrooms by school personnel. Participation was voluntary. The questionnaire was completed anonymously in the classroom and returned in a sealed envelope.

Instrument

The PSP-scale was constructed in 1987 and 1988 as a part of a questionnaire intended to address different aspects of social conditions and health among adolescents. The questionnaire covers a wide range of topics, including perceived health, health-related behaviours, school conditions, leisure activities etc. The scale construction and development was partly influenced by existing questionnaires for self-rated health among adults in general populations. In developing the questionnaire pilot studies were carried out in order to evaluate the questions and response categories and to provide suggestions for improvements. To construct the PSP-scale eight individual items are used: "had difficulty in concentrating", "had difficulty in sleeping", "suffered from headaches", "suffered from stomach aches", "felt tense", "had little appetite", "felt sad" and "felt giddy". The response categories for all of these items, which are in the form of questions, are "never", "seldom", "sometimes", "often" and "always". The five categories are ordered in terms of implied frequency and the greater the frequency, the lower the well being. The time frame concerns the school year.

Analysis

The psychometric analyses are based on the Rasch model (Andrich, 1978; Rasch, 1960/1980) and the analyses were performed using the item analysis program RUMM2020 (Andrich, Sheridan & Luo, 2004).

The PSP-scale was examined at a general level and at a finer level of analysis. At the finer level the items were examined more closely with respect to Differential Item

Functioning (DIF) across gender, year of investigation and academic orientation (theoretical program versus non theoretical program). ANOVA of standardised residuals was used to detect possible DIF (Hagquist & Andrich, 2004b). The tests of fit were carried out with a sample size adjusted to the value of the order of 2297, which is the effective sample size for test of fit within the smallest subset of data representing one single year of investigation.

In a second step items showing DIF was resolved for DIF. The lack of invariance across genders (=DIF) was assumed to reflect quantitative differences in item functioning given a unidimensional latent trait. Therefore, DIF was accounted for by splitting each misfitting item into two gender-specific items, i.e. one item for boys and one item for girls. Items were resolved for DIF step by step, starting with the worst fitting item.

Also, the dimensionality of the PSP-scale was examined more closely. Although examinations of invariance in general are likely to detect mulitidimensionality there are situations when the traditional test statistics used in Rasch analysis are less sensitive for violations of unidimensionality. Additional tests addressing mulitidimensionality specifically are therefore advisable, if violations of unidimensionality are suspected. In the present study unidimensionality is specifically examined and tested for in the following ways:

- Principal component analysis of the residuals
- Equating tests for subsets of items

Results

Table I shows the frequency distribution for the original set of eight items for the first year and the last year of investigation respectively. The table shows that the proportion of students reporting frequent complaints are higher 2005 compared to 1988, which applies to both the "often" and the "always" response categories.

Insert Table I Here

Figure 1 shows the targeting of the person-item distributions, i.e. the locations of the estimates of the item threshold parameters relative to the distribution of the estimates of the person parameters for the original set of eight items. The person locations are positively skewed with a relatively negative mean value, which is reflecting that the targeted general population of adolescents as a whole are showing a good health.

Insert Figure 1 Here

Table II shows the estimates of the item parameters and the estimates of the threshold parameters for the eight original items. The spread of the item location values means that that the items are representing different levels of severity with respect to health complaints. The item Concentrating difficulties is reflecting the lightest complaint and the items Giddy the most severe. All thresholds appear in their right order, i.e. there are

no disordered thresholds. The value of the person separation index for the original set of eight items is 0.843.

Insert Table II Here

Figure 2 shows the category probability curve for the item Felt sad. The estimates of the thresholds defining the successive categories are ordered as required. Students having a low (negative) value on the psychosomatic health scale have high probability of scoring on the lowest value on the item while students having a high (positive) value have a high probability of scoring a high value on the single item.

Insert Figure 2 Here

Table III shows analysis of variance based on standardised residuals for the original set of eight items. The table shows that all eight items work invariantly across the latent trait, i.e. there are no main effects as regards the class intervals. In contrast five out of eight items are showing gender-DIF and one item is showing DIF with respect to academic orientation.

Insert Table III Here

Figure 3 shows a graphical comparisons between boys and girls for item Felt sad showing gender-DIF.

Insert Figure 3 Here

The figure shows that the item Felt sad does not work invariantly across gender. Across the whole latent trait the observed scores for boys are located below the ICC-curve and the scores for the girls are located above the curve. The curves for boys and girls are parallel indicating a uniform DIF.

Table IV shows analysis of variance based on standardised residuals for the item set resolved for gender-DIF. The table shows that all eight items work invariantly across the latent trait, i.e. there are no main effects as regards the class intervals. Having resolved four items for DIF the scale as whole works fine. There are no or only small signs of DIF across the gender, year of investigation and academic orientation. The value of the person separation index for the set of items resolved for DIF is 0.834.

Insert Table IV Here

Testing for multidimensionality

Principal component analysis of the residuals was carried out in order to detect eventual item intercorrelation that is not accounted for by the latent trait. The analysis showed that the range of the eigenvalues for the principal components was small, indicating lack of residual correlations. This interpretation is further confirmed by the matrix for the correlations between the principal components and the eight observed items, showing that each item loads highly on only one principal component. However, the residual correlations between the first principal component and the eight items, motivates a closer examination. Since the loadings seem to direct the items into two set of items, equating tests for these two sets of items were carried out. For each person the location values for set1 and set 2 were compared and the differences assessed with independent t-tests. The outcomes from these tests confirmed lack of multidimensionality, since the location values were significantly different only for a very small number of persons (at 0.05 4.27%; at 0.01 0.8%).

Conclusions

The psychometric analysis of the PSP-scale clearly indicates that the scale is appropriate for measurement of psychosomatic health in general populations of adolescents. Importantly for such purposes, the targeting is good although there is a minor dislocation of the items versus the persons. Also, the reliability of the PSP-scale is good. The results show that the PSP-scale adequately meets measurement criteria of invariance and proper empirical ordering of the data. While the properties of the PSP-scale were fine on a general level of analysis, the DIF-analysis indicated room for improvements. Five out of eight items showed uniform DIF for gender in the initial ANOVA-analysis, i.e. these items worked differently for boys and girls. In all, four items had to be resolved for DIF in order to achieve at set of items showing invariance across genders, that is no DIF.

Since the items showing gender-DIF were favouring girls, the scale was significantly improved by taking DIF into account enabling invariant comparisons of psychosomatic problems between boys and girls. Furthermore, resolving for DIF instead of removing the DIF-items implied almost no loss of precision of measurement.

The analysis of the dimensionality of the PSP-scale indicates that the eight items can be summarised into one single scale. This is consistent with a Rasch-analysis carried out on the HBSC-instrument, but contradicts the outcomes from factor analyses on the HBSC-instrument.

Methodological remarks

Notably, although the initial DIF-analysis indicated DIF for five items, one item less had to be resolved for DIF by gender. In addition, only two of these four items showed DIF in the initial analysis. The items showing DIF across genders were split into gender-specific items. The procedure carried out confirms that items showing DIF should not be resolved for simultaneously in one step but consecutively. It also demonstrates the relative nature of DIF, i.e. the operating characteristics of an item are not absolute but conditioned on which other items are included in the item set.

References

Achenbach, T.M. (1991a). Manual for the Youth Self Report and 1991 Profile. Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T.M. (1991b). Manual for the Child Behavior Checklist/4-18 and 1991 Profile. Burlington, VT: University of Vermont Department of Psychiatry.

Andrich, D. (1978). A rating formulation for ordered response categories. Psychometrika, 43, 561-573.

Andrich, D., Sheridan, B. & Luo, G. (2004). RUMM2020: A Windows interactive program for analysising data with Rasch Unidimensional Models for Measurement. RUMM Laboratory, Perth, Western Australia.

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. Journal of the American Academy of Child and Adolescent Psychiatry, 40, 1337-1345.

Hagquist, C. & Andrich, D. (2004a). Measuring subjective health among adolescents in Sweden: A Rasch-analysis of the HBSC-Instrument. Social Indicators Research, 68, 201-220.

Hagquist, C. & Andrich, D. (2004b). Is the Sense of Coherence-Instrument applicable on adolescents? A latent trait analysis using Rasch-modelling. Personality and Individual Differences, 36, 955-968.

Rasch, G. (1960/80). Probabilistic models for some intelligence and attainment tests. (Copenhagen, Danish Institute for Educational Research). Expanded edition (1980) with foreword and afterword by B.D. Wright, (1980). Chicago: The University of Chicago Press.

Table I. The proportion of responses in different categories for eight items 1988 and 2005.

	Response category									
	Never (0)		Seldom (1)		Sometimes (2)		Often (3)		Always (4)	
Item label	**1988**	**2005**	**1988**	**2005**	**1988**	**2005**	**1988**	**2005**	**1988**	**2005**
Concentrating difficulties	5.3	7.4	23.1	24.4	51.7	41.2	18.3	21.8	1.6	5.0
Sleeping difficulties	20.6	19.6	29.4	28.3	31.8	28.5	15.4	16.4	2.8	6.7
Headache	24.3	18.3	31.7	29.2	29.0	28.9	13.6	17.4	1.4	5.8
Stomach ache	29.9	27.8	37.3	32.3	23.1	23.8	8.7	12.0	1.0	3.3
Felt tense	19.3	20.1	38.6	31.9	31.9	28.0	8.8	15.0	1.1	4.3
Little appetite	34.7	37.9	33.8	28.6	22.8	19.2	6.8	9.5	1.7	4.0
Felt sad	13.8	19.7	35.5	29.1	37.6	28.2	12.0	17.6	0.9	4.6
Giddy	39.5	32.7	30.9	31.7	21.2	21.0	7.5	10.6	0.7	3.3

Table II. Estimates of individual item parameters and threshold parameters in the original set of 8 items. Chi^2 test based on original sample size and adjusted sample size respectively.

Item label	**Location Estimate**	**P-val Chi2-test**		**Thresholds**			
		n= 14895	n= 2297	1	2	3	4
Concentrating difficulties	-0.730	0.000	0.340	-2.576	-1.088	0.946	2.718
Sleeping difficulties	-0.159	0.000	0.795	-1.562	-0.642	0.503	1.701
Headache	-0.191	0.000	0.786	-1.744	-0.627	0.322	2.050
Stomach ache	0.322	0.000	0.087	-1.799	-0.623	0.352	2.070
Felt tense	0.043	0.000	0.011	-2.093	-0.620	0.689	2.024
Little appetite	0.294	0.000	0.029	-1.315	-0.450	0.472	1.292
Felt sad	-0.081	0.000	0.567	-2.047	-0.756	0.527	2.276
Giddy	0.502	0.000	0.036	-1.558	-0.741	0.263	2.036

Table III. Analysis of variance of residuals for test of DIF between gender, grades and years of investigations as well as tests of class interval fit. Number of class intervals=10. Adjusted sample size for test of fit (n=2297).

	Probability values								
	Division by gender			Division by years of investigations			Division by program		
	Class interval	Gender	Gender by class interval	Class interval	Year	Year by class interval	Class interval	Program	Program by class interval
Item label									
Concentrating difficulties	0.281	0.000	N/Sig	0.321	0.019	1.000	0.429	0.000	0.993
Sleeping difficulties	0.778	0.000	N/Sig	0.803	0.034	1.000	0.933	0.646	0.987
Headache	0.731	0.009	N/Sig	0.748	0.009	1.000	0.817	0.554	1.000
Stomach ache	0.058	0.000	N/Sig	0.072	0.963	1.000	0.435	0.382	1.000
Felt tense	0.003	0.525	0.467	0.003	0.280	1.000	0.012	0.412	0.951
Little appetite	0.033	0.036	0.999	0.038	0.371	1.000	0.271	0.352	1.000
Felt low	0.441	0.000	N/Sig	0.467	0.732	1.000	0.357	0.000	0.890
Giddy	0.036	0.000	0.018	0.039	0.842	1.000	0.556	0.296	0.987

Table IV. Analysis of variance of residuals for test of DIF between genders, grades and years of investigations as well as tests of class interval fit. Number of class intervals=10. Adjusted sample size for test of fit (n=2297).

	Probability values								
	Division by gender			Division by years of investigations			Division by program		
Item label	Class interval	Gender	Gender by class interval	Class interval	Year	Year by class interval	Class interval	Program	Program by class interval
Concentrate	0.496	0.563	0.982	0.522	0.018	1.000	0.563	0.002	0.999
Sleeping	0.835	0.563	1.000	0.846	0.032	1.000	0.971	0.349	1.000
Appetite	0.047	0.020	0.940	0.047	0.386	1.000	0.343	0.619	1.000
Giddy	0.002	0.660	0.991	0.003	0.849	1.000	0.237	0.553	0.997
Sad Boys	0.991	N/A	N/A	0.992	0.242	1.000	0.999	0.004	0.999
Sad Girls	0.880	N/A	N/A	0.893	0.935	1.000	0.796	0.513	0.987
Stomach Boys	0.451	N/A	N/A	0.483	0.982	1.000	0.806	0.252	0.992
Stomach Girls	0.936	N/A	N/A	0.944	0.951	1.000	0.999	0.270	1.000
Headache Boys	0.986	N/A	N/A	0.988	0.139	1.000	0.986	0.362	0.985
Headache Girls	0.991	N/A	N/A	0.992	0.102	1.000	0.986	0.881	0.999
Tense Boys	0.323	N/A	N/A	0.355	0.976	1.000	0.384	0.145	0.991
Tense Girls	0.075	N/A	N/A	0.087	0.062	1.000	0.313	0.135	0.972

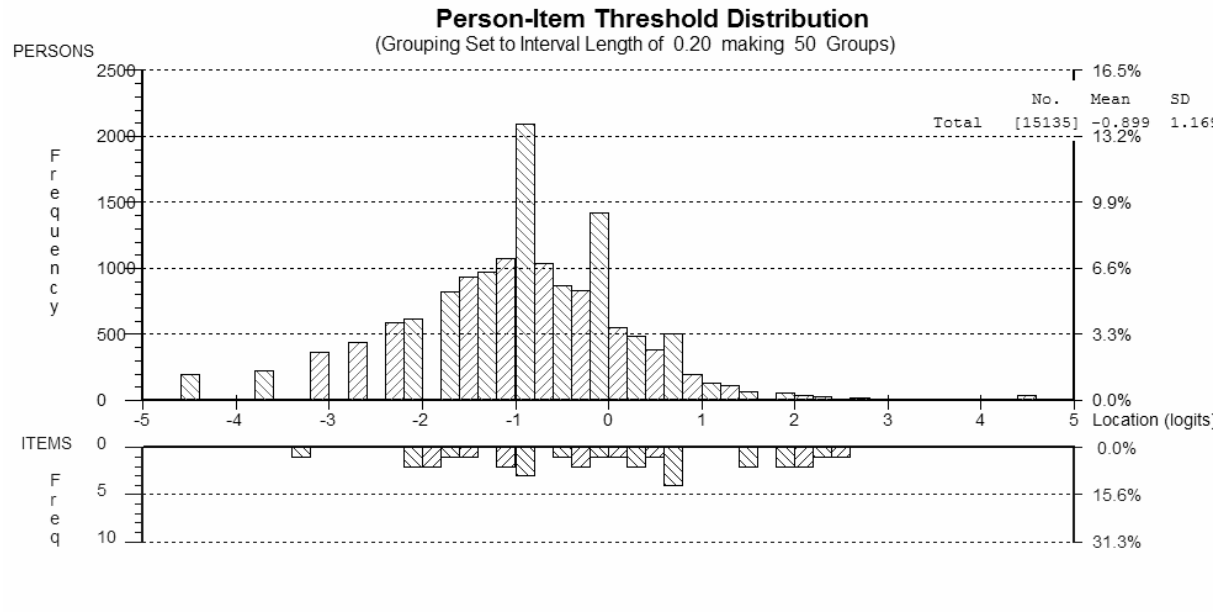

Figure 1. Person-item threshold distribution. The higher the score, the worse health.

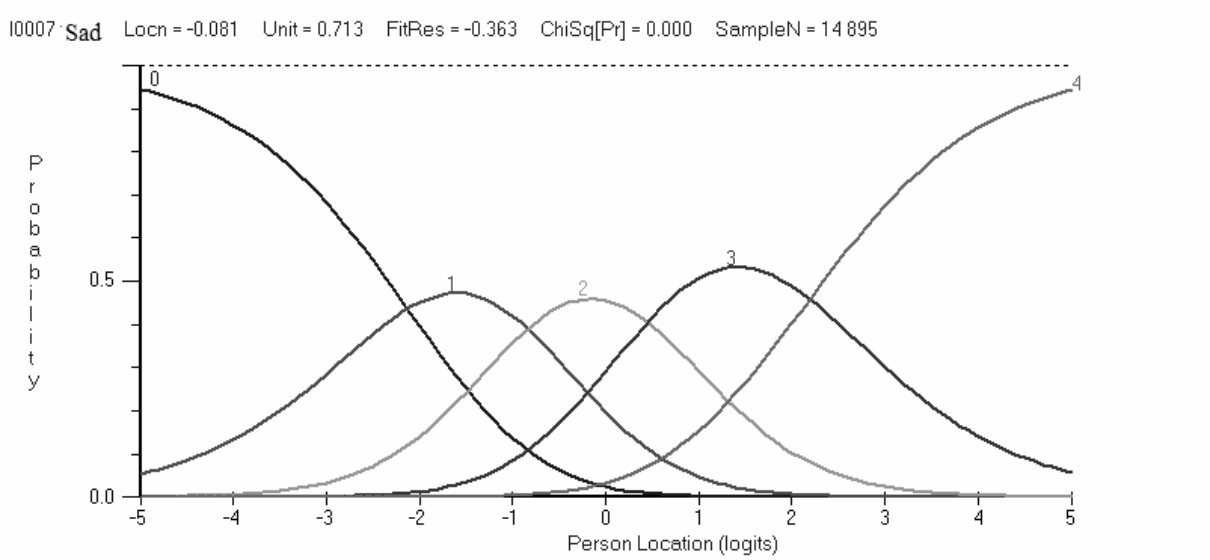

Figure 2. Category Probability Curve for item Felt sad.

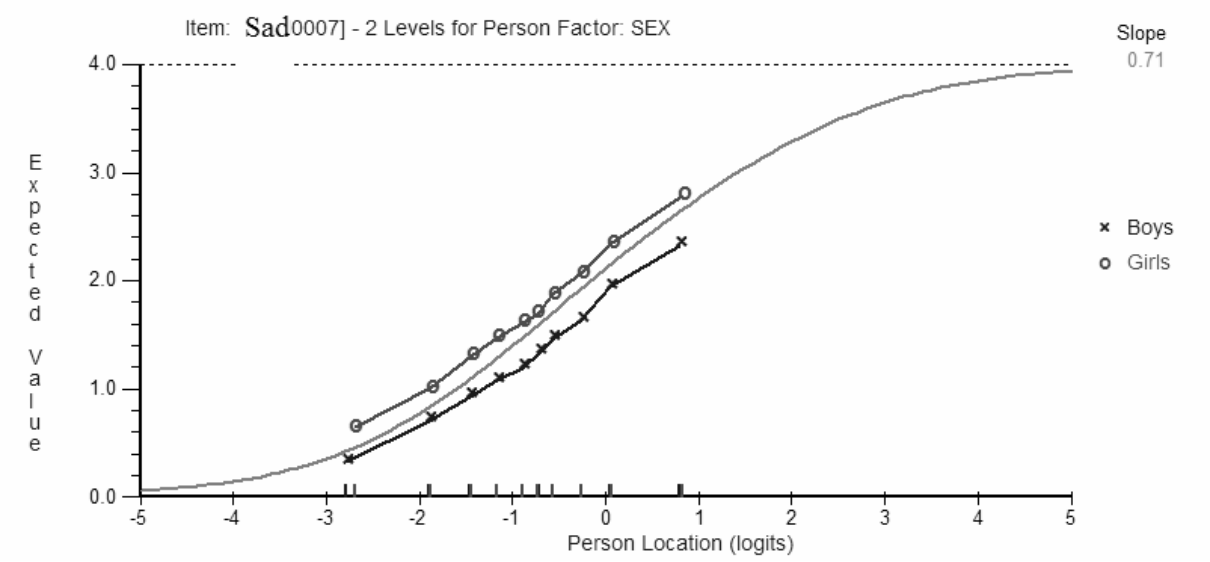

Figure 3. Item Characteristic Curve for item Felt sad.