Current Psychology, 2019
https://doi.org/10.1007/s12144-019-00424-5

Rasch analysis of the Satisfaction with Life Scale across countries:

Findings from South Africa and Italy

Lusilda Schutte[1*], Luca Negri[2], Antonella Delle Fave[3], & Marié P. Wissing[4]

[1] *Africa Unit for Transdisciplinary Health Research (AUTHeR), North-West University, South Africa. E-mail: Lusilda.Schutte@nwu.ac.za. Orcid ID: 0000-0002-2107-3669*

[2] *Department of Pathophysiology and Transplantation, University of Milano, Italy. Email: luca.negri@unimi.it.*

[3] *Department of Pathophysiology and Transplantation, University of Milano, Italy. Email: antonella.dellefave@unimi.it. Orcid ID: 0000-0003-1381-6538*

[4] *Africa Unit for Transdisciplinary Health Research (AUTHeR), North-West University, South Africa. E-mail: Marie.Wissing@nwu.ac.za. Orcid ID: 0000-0002-0445-6246*

[*]Corresponding author. Address: Africa Unit for Transdisciplinary Health Research (AUTHeR), North-West University, Potchefstroom Campus, 11 Hoffman Street, Potchefstroom, 2531. E-mail: Lusilda.Schutte@nwu.ac.za. Telephone number: +27 18 299 1104.

Acknowledgements

The South African research was financially supported in part by the National Research Foundation of South Africa (grant number: 106050). The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed are that of the authors, and that the NRF accepts no liability whatsoever in this regard.

Abstract

The Satisfaction with Life Scale (SWLS) is widely used to assess global satisfaction with life. Although the scale's psychometric properties were extensively investigated through different methods, Rasch analysis can provide further insight into the scale's targeting and sensitivity across the different levels of the underlying construct. It also allows researchers to evaluate how well a scale fits the assumptions of unidimensionality and local independence of the items, how the response categories are used, and whether differential item functioning occur in different demographic groups. In this study, Rasch analysis was used to examine the psychometric properties of the SWLS among participants from South Africa (n = 676) and Italy (n = 516). Findings showed that the scale was insensitive at high levels of life satisfaction. Since the majority of the participants reported high scores, these results suggest that the SWLS may not be sensitive to detect change in the general population. Although support was provided to the scale's unidimensional factor structure, a distinction emerged between items referring to satisfaction with present and past life. No group difference in item functioning was detected for country, gender, age group, or education level. However, findings suggest that using fewer response categories with less nuanced lower level descriptors could be more appropriate, especially for the South African sample. Overall, results highlight the need for further research on the SWLS, especially concerning insensitivity at the upper range.

Keywords: Satisfaction with Life Scale (SWLS), Rasch analysis, sensitivity and targeting, dimensionality, response category functioning, differential item functioning

Rasch analysis of the Satisfaction with Life Scale across countries:

Findings from South Africa and Italy

The Satisfaction with Life Scale (SWLS) (Diener et al. 1985) is one of the instruments most commonly used to assess the cognitive component of subjective well-being. Satisfaction with life is often considered as an indicator of individual and societal progress, and its assessment is deemed as useful to evaluate the impact of individual and group-level interventions and to orient public policy development (Diener et al. 2018; Diener and Seligman 2004; Diener and Tay 2016; Stephen et al. 2013). Considering the relevance attributed to life satisfaction across disciplines and social contexts, it is important to verify whether instruments used to measure it are reliable, valid and sensitive to variation and change at all levels of the latent construct.

Although the internal consistency, reliability and validity of the SWLS were extensively supported (Diener et al. 1985; Pavot and Diener 1993, 2008; Vassar 2008), important issues were raised about the scale's dimensionality. While some studies supported the initially intended one-factor structure (e.g., Lorenzo-Seva et al. 2019; Ortuño-Sierra et al. 2019; López-Ortega et al. 2016), other ones suggested the presence of a temporal dimension, requiring the distinction between satisfaction with the present and past life (e.g., Bai et al. 2011; Clench-Aas et al. 2011). In addition, investigating the sensitivity of the SWLS along the latent trait continuum is of fundamental importance, considering that the scale is often used to evaluate change, impact, and progress on individual and societal levels. To the best of our knowledge, the only study that explored the scale's sensitivity was conducted by Vittersø et al. (2005). Findings showed that the scale had varying levels of sensitivity and discriminatory power for different latent classes of respondents from Norway and Greenland. Controversial findings were also obtained as concerns measurement invariance across samples differing in culture, gender, and age (see Emerson et al. 2017 for a review). The lack of consensus regarding the psychometric properties of SWLS calls for further research into this matter.

A very suitable analytical approach to address these issues is Rasch analysis, which is particularly appropriate for (intended) one-dimensional instruments, such as the SWLS. Rasch analysis can provide insight into a scale's targeting and sensitivity across the different levels of the underlying construct. It also

allows researchers to evaluate how well a scale fits the assumptions of unidimensionality and local independence of the items, how the response categories are used, and whether differential item functioning occur in different demographic groups (Bond and Fox 2007). Vittersø et al. (2005) fitted the SWLS to a mixed Rasch model (an extension of the Polytomous Rasch Model that allows threshold parameters to vary across groups, Rost and Langeheine 1997), but they pursued the specific aim of examining group differences in response patterns comparing data collected in Norway and Greenland. A more general account of how the SWLS fits the Rasch model that considers, inter alia, the scale's dimensionality, sensitivity, response category functioning, and differential item functioning, will shed further light on the functioning of the scale. Therefore, the aim of this study was to explore the psychometric properties of the SWLS for samples from South Africa and Italy using a Rasch modelling framework.

Method

Participants and Procedure

Using a cross-sectional survey design, convenience samples were gathered from 1192 adult participants from South Africa (n = 676) and Italy (n = 516), who completed the English and Italian versions of the SWLS, respectively. In South Africa, participants were recruited by word-of-mouth, where already included participants referred other potential participants. In Italy, multiple sources of participants were contacted (e.g., companies, schools, laboratories, personal acquaintances). Inclusion criteria comprised being 18 years or older, having attained at least secondary education, and, for South Africans, being fluent in English. Participants' demographic features are reported in Table 1. Participation was voluntary and followed written informed consent. Questionnaires were completed in paper-and-pencil format. The regulatory ethics committees in the two countries granted ethical approval.

Measures

Satisfaction with Life Scale (Diener et al. 1985). This instrument measures satisfaction with life as a whole on a subjective cognitive-judgemental level, through five scaled items on a 7-point rating scale ranging from 1 = strongly disagree to 7 = strongly agree. As concerns content formulation, items 1 to 3 refer to satisfaction with the present, and items 4 and 5 to satisfaction with the past. The scale exhibited reliability and validity in various contexts and cultures (Pavot and Diener 1993, 2008). When the English version of

the scale was administered to a multicultural South African sample, confirmatory factor analysis suggested poor fit (root mean square error of approximation [RMSEA] = 0.96) with Cronbach's alpha values of .86 and .70 for white and black samples, respectively (Wissing and Van Eeden 2002). In a sample of Italian adolescents and young adults, the Italian version displayed marginal fit (RMSEA = .103 and comparative fit index [CFI] = .975) with a Cronbach's alpha value of .85 (Di Fabio and Gori 2016).

Data Analysis

The Rasch Rating Scale Model (Andrich 1978) was applied using Winsteps® 3.81 (Linacre 2014b). In the ability testing context where Rasch modelling was developed, the level of the underlying construct captured by an item is labelled as 'item difficulty', whereas 'person ability' refers to the respondent's level in the underlying construct. In the current study, considering the SWLS aim and contents, we deemed as more appropriate to refer to 'item challenge' and 'person intensity', respectively. R Version 3.3.2 was used to draw the density functions of the item category challenge and person intensity levels obtained from Winsteps® 3.81. Data were initially analysed for South Africa and Italy separately; the dataset was subsequently combined to assess differential item functioning. Different aspects of the SWLS were explored through Rasch analysis.

Sensitivity: Targeting and order of item challenge levels. Rasch analysis can be used to check how well a scale targets the sample at hand (Chao and Green 2013). The output from Winsteps® 3.81 was graphically represented using R Version 3.3.2. Density functions of the average challenge levels of the items' response categories and participants' average intensity levels were reported on the same graph. The person-item maps were used to inspect the person and item distribution, and to examine the order of the items in terms of their challenge levels.

Person and item separation and reliability. Person separation and reliability indices indicate the level of distinction that can be detected among persons along the measured variable (Bond and Fox 2007). Values larger than 2 and 0.8, respectively, imply that two categories of persons - high versus low scorers - can be distinguished (Linacre 2014a). Item separation and reliability indices indicate the extent to which item progression would be stable across samples (Bond and Fox 2007). Values larger than 3 and 0.9, respectively, confirm the item challenge order on three levels of item challenge (Linacre 2014a).

Item Fit, Unidimensionality, and Local Independence. The Rasch model prescribes that useful measurement involves locally independent, monotonically increasing items that measure a unidimensional construct. When using the rating scale model, item infit or outfit mean square statistics lower than 0.6 or higher than 1.4 indicate overfit and underfit, respectively (Bond and Fox 2007). Positive point-biserial correlations indicate that item scores increase at increasing levels of the underlying construct (Linacre 2014a). Rasch principal components analysis (PCA) of the residuals gives an indication of unidimensionality when a small eigenvalue of the first contrast (e.g., < 2.0) is obtained, and a large proportion variance (e.g., > 40%) is explained by the Rasch component. If the loadings of the item residuals on the first contrast display contrasting patterns, multidimensionality may be present (Linacre 2014a). Small item residual pair correlations (e.g., around 0.4) support local independence (Linacre 2014a).

Response category functioning. Rasch analysis provides information about the pattern of scale use. The category threshold estimates should increase monotonically, each response category should contain at least some observations and represent a distinct portion of the latent trait, and response categories' infit and outfit mean square statistics should be lower than 2.0 (Bond and Fox 2007). By exploring the model fit after collapsing response categories, researchers can get suggestions on how to adapt the rating scale, if necessary (Fox and Jones 1998; Tennant and Conaghan 2007).

Differential item functioning (DIF). DIF occurs when people with equal levels of the latent construct respond differentially to an item. In the current study, the degree of uniform DIF (Tennant and Conaghan 2007) was assessed for country, gender, education level, and age group by means of the DIF Contrast, where values ≥ 0.64 indicate moderate to large DIF (Linacre 2014a). Due to the sensitivity of Mantel-Haenszel statistic to large sample sizes, DIF contrasts were deemed as adequate to interpret the findings. For sake of completeness, however, the polytomous version of the Mantel-Haenszel statistic with its *p*-value (Mantel 1963; Mantel and Haenszel 1959), and the Bonferroni-corrected 5% significance level against which *p*-values should be compared for each demographic variable are also reported in the results section.

Results

Targeting and order of item challenge levels

Figure 1 shows the person-item threshold distribution. The levels of life satisfaction measured by the items (i.e., item challenge levels) were lower than the intensity levels attained by the majority of the persons. In other words, most participants reported levels of life satisfaction for which the scale did not contain much information and would therefore not be sensitive to detect variation or change.

Table 2 shows the item challenge levels with their standard errors, infit and outfit mean square values, point-biserial correlations, and item residual loadings on the first contrast of the Rasch PCA of residuals. In the Electronic Supplemental Material (ESM), Figure 1S, the person-item threshold maps is displayed. Item 5 ('If I could live my life over, I would change almost nothing') was the most challenging item to endorse for South African participants, and, together with item 2 ('The conditions of my life are excellent'), the hardest to endorse for Italians. For both samples, the least challenging item to endorse was item 3 ('I am satisfied with my life').

Person and item separation and reliability

The person separation and reliability indices were 1.93 and .79, respectively, for South Africa, and 2.55 and .87 for Italy. The item separation and reliability indices were 5.00 and .96, respectively, for South Africa and 6.22 and .97 for Italy.

Item Fit, Unidimensionality and Local Independence

As reported in Table 2, item 5 did not adequately fit the Rasch model in both samples. Removal of this item yielded a model where item 4 ('So far I have gotten the important things I want in life') displayed misfit (infit = 1.46 and outfit = 1.44 for South Africa; infit = 1.47 and outfit = 1.43 for Italy). After removal of both items 5 and 4, all remaining items fitted the Rasch model, suggesting that items 5 and 4 are potentially responsible for deviations from the model's unidimensionality assumption. All point-biserial correlations were positive for both groups.

In the Rasch PCA of residuals, the Rasch component explained 60.2% of the variance for the South African sample and 69.3% for the Italian sample; the eigenvalue of the first contrast was 1.8 for both groups. In the standardised residual plot of the first contrast (ESM Figure 2S), items 1 to 3 clustered together, while

items 4 and 5 formed a second cluster for both countries. Although this clustering pattern suggests a two-dimensional structure of the SWLS, other dimensionality indicators such as the eigenvalue of the first contrast and the percentage variance explained by the first component were within acceptable limits, suggesting that the single-factor model was also appropriate. All item residual pair correlations were small (≤.06 for South Africa and ≤.10 for Italy), supporting local independence.

Response category functioning

Figure 2 shows two category probability curves for each country, one for the standard SWLS (left side) and one with collapsed response categories (right side). ESM Table 1S presents the rating scale functioning. As concerns South Africa, the standard curve shows that categories 2 and 3 were the most likely categories to be endorsed for a small portion of the life satisfaction continuum, which suggests that they may have been redundant. Response category 1's outfit statistic was also not within allowable limits (ESM Table 1S). For Italy, despite the overall better category functioning in the standard curve, category 3 was the most likely category to be endorsed for a small section of the latent trait continuum. After collapsing the lower response categories (1 to 3) an improved picture emerged for both countries (Figure 2, right side), with each response category representing a clearly distinct portion of the underlying trait.

Differential item functioning

DIF for country, gender, and education level is presented in Table 3. The age group combinations displaying minimum and maximum DIF are reported in Table 4. Although the *p*-value of the Mantel-Haenszel statistic in comparison with the Bonferroni-corrected significance level suggests statistically significant DIF for country on item 2 and for gender on items 1 and 4, in all these cases the DIF contrasts were smaller than the 0.64 guideline (Linacre 2014a). No statistically significant DIF was detected for education level or age group. Altogether, considering the DIF contrasts, no DIF was detected for country, gender, education level, or age group.

Discussion

This study aimed to explore the psychometric properties of the SWLS with samples from South Africa and Italy using a Rasch modelling framework. Findings showed that the scale was not sensitive at high levels of the construct, but also that most participants fell in that range. The unidimensionality of the

SWLS was confirmed in both samples, although a distinction was detected between items assessing satisfaction with present and past life. The number of response categories seemed to be excessive, particularly for the South African sample, suggesting the usefulness of exploring less nuanced descriptors for lower categories. No DIF was detected for country, gender, age group, or education level.

Sensitivity and Targeting

A high density of item challenge thresholds was detected for low to moderate levels of life satisfaction; towards higher levels of life satisfaction the threshold density decreased, whereas the person intensity levels peaked. This means that the scale provided maximum information / had maximal discriminatory power at life satisfaction levels lower than the levels attained by most participants. These findings suggest that the SWLS might be highly sensitive for variation and change at low and moderate levels of satisfaction, but less sensitive (i.e., provide less information) at the upper range of the latent trait.

Our findings are substantially consistent with those obtained by Vittersø et al. (2005). These authors detected varying levels of SWLS sensitivity for different latent classes of participants. The scale did not discriminate well between individuals with high versus low life satisfaction within the most frequent class (39.1%, 80.9%, and 22.4% of the pooled, Greenlandic, and Norwegian samples, respectively), comprising participants who tended to select extreme response options and displayed a large degree of random responding. Better discrimination was obtained for the class ranking second in frequency (37.0%, 11.1%, and 47.3% of the pooled, Greenlandic, and Norwegian samples, respectively), grouping the participants who tended to avoid both high and low extreme scores. However, the person parameter plot reported in Vittersø et al. (2005) for this second class suggests that the discriminatory power of the SWLS was relatively large for high versus low life satisfaction, while for varying levels of high life satisfaction – characterising most participants in this class – the gradient was rather flat. The SWLS discriminated well all along the latent trait continuum for the third most frequent class (21.8%, 7.4%, and 27.6% of the pooled, Greenlandic, and Norwegian samples, respectively), that included participants who did not consider their life conditions to be excellent, but at the same time tending to agree with item 5 ("*If I could live my life over, I would change almost nothing*"). Notably, a fairly normal distribution of scores centred around the midpoint characterised

this class, a peculiar pattern vis-à-vis the negatively skewed distribution detected for the two more frequent latent classes.

Our findings are consistent with those of Vittersø et al. (2005), suggesting that altogether the SWLS does not discriminate well between life satisfaction scores on the upper end of the scale, where the ratings of most participants across countries are grouped (e.g., Hinz et al. 2018; Di Fabio and Gori 2016; Jovanović 2016; Clench-Aas et al. 2011). These results apparently contradict a review (Diener et al. 2013) highlighting that life satisfaction scales reflect the changes in satisfaction with life expected after events such as unemployment, childbirth, psychotherapy, dementia of spousal partners, assault, disability, and changes in marital status. It is however plausible that such major life events require the scale to be sensitive for change at lower ranges of life satisfaction. Moreover, a decline in score precision at higher levels of the scale was also detected for other well-being measures, as reported by Schutte et al. (2016), who performed a Rasch analysis on the Meaning in Life Questionnaire – Presence subscale (MLQ-P) (Steger et al. 2006), and by Abbott et al. (2010) who used normal ogive item response theory to analyse data collected with the Psychological Well-being Scales (PWBS) (Ryff 1989).

The inability to detect change at upper levels of life satisfaction is significant, as the SWLS – like other well-being measures – was developed to complement scales measuring psychopathology by specifically targeting positive mental health components in the general population. Since the majority of individuals across countries typically report moderate to high levels of life satisfaction (e.g., Hinz et al. 2018; Di Fabio and Gori 2016; Jovanović 2016; Clench-Aas et al. 2011), the inadequate sensitivity of the SWLS along the full spectrum of scores casts doubts on the appropriateness of its use in epidemiological studies and in studies of impact and progress. Findings in this regard also have implications for fields such as counselling and positive psychology, where the effectiveness of interventions among non-clinical samples are often measured in terms of increase in satisfaction with life (e.g., Berger et al. 2019; Kees and Rosenblum 2015; Proyer et al. 2013), disregarding the fact that at baseline a large proportion of participants already report moderate to high levels of satisfaction with life.

The lack of sensitivity of the SWLS at upper levels can be interpreted in different ways, leading to different strategies for dealing with the matter. Sensitivity problems could be related to a suboptimal

formulation of instructions, items, and response options that could be revised in order to better target high levels of life satisfaction. To this purpose, the effect of using fewer lower response category descriptors could be investigated. The lack of sensitivity may also be ascribed to the classical test theory approach that was applied in the development and initial validation of the SWLS, where the aim is to minimize floor and ceiling effects (i.e., the proportion of participants with minimum and maximum scores) (Petrillo et al. 2015). Using an item response theory approach to develop and/or adapt satisfaction with life measures may address this issue, since it specifically aims to provide information across the entire intended score range (Petrillo et al. 2015). According to a further interpretation, life satisfaction could be considered a quasi-trait (Reise and Waller 2009), containing variation only at the low end of the latent trait continuum.

The present study also revealed that item 3 ('I am satisfied with my life') was the least challenging item for both samples; this finding was consistent with that obtained among Norwegian and Greenlandic participants (Vittersø et al. 2005). Single-item life satisfaction measures similar to item 3 are often used in large-scale nationally representative panel studies, population studies, and international studies on the determinants of quality of life (Fujita and Diener 2005; Jovanović 2016). The fact that this item tends to be relatively easy to endorse suggests that it describes a condition suitable even to people with lower levels of life satisfaction. This implies that its use as a single-item may further reduce the sensitivity of the life satisfaction assessment at the upper end of the continuum. In this regard, Michalos and Kahlke (2010) recommend the use of multiple measures of perceived quality of life, rather than just a single item measuring life satisfaction.

Factor structure of the SWLS

In line with previous studies (Diener et al. 1985; Pavot and Diener 1993; López-Ortega et al. 2016; Ortuño-Sierra et al. 2019; Lorenzo-Seva et al. 2019), our findings provide support to the unidimensional factor structure of the SWLS in two different countries, based on standard guidelines for the interpretation of results from Rasch analysis (cf., Bond and Fox 2007). At the same time, a temporal dimension of life satisfaction was identified, by distinguishing two factors referring to satisfaction with the present (items 1 to 3) and past life (items 4 and 5), respectively. This pattern was also identified in other studies (McDonald 1999; Bai et al. 2011; Clench-Aas et al. 2011; Jovanović 2017) and different statistical models were applied

to account for this pattern. For example, McDonald (1999) fitted a hierarchical factor model where the two temporal dimensions were incorporated into a general life satisfaction factor, while Bai et al. (2011) fitted a one-factor model with a wording-effect on items 4 and 5. Clench-Aas et al. (2011) and Jovanović (2017) applied a modified one-factor model with a residual correlation between items 4 and 5.

Response category functioning

The findings from this study suggest that, particularly for the South African sample, seven response categories were too many, with redundancy emerging at lower levels of life satisfaction. The use of fewer response categories referring to less nuanced lower category descriptors (e.g., 1 = *Disagree*, 2 = *Slightly agree*, 3 = *Moderately agree*, 4 = *Mostly agree*, and 5 = *Completely agree*) could help address this issue and should be explored in future. Problematic functioning of a balanced 7-point response scale at low levels was also detected in a Rasch analysis of data collected with the Meaning in Life Questionnaire in South Africa, Australia, and New Zealand (Schutte et al. 2016). The differences detected between the South African and Italian samples may be ascribed to cultural factors, in line with previous studies showing that rating scale use of life satisfaction measures differs across cultural groups (Diener et al. 2013; Vittersø et al. 2005).

Differential item functioning

Based on the DIF contrasts, the SWLS did not exhibit differential item functioning when country, gender, age group, or education levels were taken into account. This finding is consistent with the broad and multifaceted literature on the measurement invariance of the SWLS (see Emerson et al. 2017 for a review), supporting the scale's potential for use across groups differing in demographic variables such as culture, gender, education level and age.

Study limitations and future directions

Although this study presents novel findings, it is not without limitations. SWLS data were fitted to the Rasch model, which is a simple IRT model with only item difficulty as parameter. Despite the attractive mathematical properties of the Rasch model, more complex IRT approaches may shed further light on a scale's performance. In addition, the current study did not consider different latent classes of participants, that were taken into account in the mixed Rasch model adopted by Vittersø et al. (2005), leading to the emergence of related differences in sensitivity and response styles. The possibility that response category

use can differ across latent classes within the same sample and cultural group needs to be further explored in future research.

While the present study was focused on convenience samples from two countries, replication in other cultural groups and in representative samples would be welcome. Also measurement invariance could be investigated through different models, such as multigroup confirmatory factor analysis. Additional efforts are required to identify the best model to represent the higher order unidimensionality of the SWLS, while taking into consideration the distinction between satisfaction with the past and present.

Altogether, a clearer theoretical understanding of the conditions of low and high life satisfaction is necessary, in order to develop instruments that reflect variation along the whole latent trait continuum (Maul 2017). In this regard, qualitative approaches could provide more in-depth insight into the conceptual nature of life satisfaction and its manifestations along the latent trait continuum, from the twofold perspective of construct structure and lay people understanding (Carlquist et al. 2017).

Conclusions

Although the SWLS is widely used and well-investigated in research and practice, the employment of Rasch analysis on data from two countries shed new light on the scale's functioning. The globally good psychometric properties of the scale were confirmed; problems with sensitivity were however detected in the scale portion referring to high levels of life satisfaction. This crucial issue needs to be adequately addressed by researchers, especially if we consider that satisfaction with life is often an outcome variable of the growing amount of interventions aimed at promoting well-being in non-clinical samples. The opportunities offered by Rasch analysis and other analytical approaches should be exploited by researchers interested in investigating the psychometric features of well-being related instruments, in order to provide professionals and clinicians with effective and reliable measurement tools.

Conflict of Interest

None of the authors have conflicting interests to declare.

References

Abbott, R. A., Ploubidis, G. B., Huppert, F. A., Kuh, D., & Croudace, T. J. (2010). An evaluation of the precision of measurement of Ryff's Psychological Well-Being Scales in a population sample. *Social Indicators Research, 97,* 357-373, doi:10.1007/s11205-009-9506-x.

Andrich, D. (1978). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement, 2*(4), 581-594, doi:0.1080/14330237.2014.904088.

Bai, X., Wu, C., Zheng, R., & Ren, X. (2011). The psychometric evaluation of the Satisfaction with Life Scale using a nationally representative sample of China. *Journal of Happiness Studies, 12*(2), 183-197, doi:10.1007/s10902-010-9186-x.

Berger, P., Bachner-Melman, R., & Lev-Ari, L. (2019). Thankful for what? The efficacy of interventions targeting interpersonal versus noninterpersonal gratitude. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 51*(1), 27-36, doi:10.1037/cbs0000114.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). New York, NY: Routledge.

Chao, R. C.-L., & Green, K. E. (2013). Rasch analysis of the Outcome Questionnaire with African Americans. *Psychological assessment, 25*(2), 568-582, doi:10.1037/a0032083.

Clench-Aas, J., Nes, R. B., Dalgard, O. S., & Aarø, L. E. (2011). Dimensionality and measurement invariance in the Satisfaction with Life Scale in Norway. *Quality of Life Research, 20,* 1307-1317, doi:10.1007/s11136-011-9859-x.

Di Fabio, A., & Gori, A. (2016). Measuring adolescent life satisfaction: Psychometric properties of the Satisfaction With Life Scale in a sample of Italian adolescents and young adults. *Psychoeducational Assessment, 34,* 501-506, doi:10.1177/0734282915621223.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49*(1), 71-75.

Diener, E., Inglehart, R., & Tay, L. (2013). Theory and validity of life satisfaction scales. *Social Indicators Research, 112,* 497-527, doi:10.1007/s11205-012-0076-y.

Diener, E., Lucas, R. E., & Oishi, S. (2018). Advances and open questions in the science of subjective well-being. *Collabra: Psychology, 4*(1), 1-49, doi:10.1525/collabra.115.

Diener, E., & Seligman, M. E. P. (2004). Beyond money: Toward an economy of well-being. *Psychological Science in the Public Interest, 5*(1), 1-31.

Diener, E., & Tay, L. (2016). New frontiers: Societal measures of subjective well-being for input to policy. In L. Bruni, & P. L. Porta (Eds.), *Handbook of research methods and applications in happiness and quality of life* (pp. 35-52). Northampton, MA, US: Edward Elgar Publishing.

Emerson, S. D., Guhn, M., & Gadermann, A. M. (2017). Measurement invariance of the satisfaction with life scale: Reviewing three decades of research. *Quality of Life Research*, doi:10.1007/s11136-017-1552-2.

Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology, 45*(1), 30, doi:0022-0167/98/53.00.

Fujita, F., & Diener, E. (2005). Life satisfaction set point: Stability and change. *Journal of personality and social psychology, 88*(1), 158-164, doi:10.1037/0022-3514.88.1.158.

Hinz, A., Conrad, I., Schroeter, M. L., Glaesmer, H., Brähler, E., Zenger, M., et al. (2018). Psychometric properties of the Satisfaction with Life Scale (SWLS), derived from a large German community sample. *Quality of Life Research, 27*, 1661-1670, doi:10.1007/s11136-018-1844-1.

Jovanović, V. (2016). The validity of the Satisfaction with Life Scale in adolescents and a comparison with single-item life satisfaction measures: A preliminary study. *Quality of Life Research, 25*(12), 3173-3180, doi:10.1007/s11136-016-1331-5.

Jovanović, V. (2017). Measurement invariance of the Serbian version of the Satisfaction With Life Scale across age, gender, and time. *European Journal of Psychological Assessment*, doi:10.1027/1015-5759/a000410.

Kees, M., & Rosenblum, K. (2015). Evaluation of a psychological health and resilience intervention for military spouses: A pilot study. *Psychological Services, 12*(3), 222-230, doi:10.1037/ser0000035.

Linacre, J. M. (2014a). *A user's guide to Winsteps Ministep Rasch-model computer programs: Program manual 3.81.0*. Beaverton, Oregon: Winsteps.com.

Linacre, J. M. (2014b). Winsteps (Version 3.81.0). Beaverton, Oregon: Winsteps.com.

López-Ortega, M., Torres-Castro, S., & Rosas-Carrasco, O. (2016). Psychometric properties of the Satisfaction with Life Scale (SWLS): Secondary analysis of the Mexican Health and Aging Study. *Health & Quality of Life Outcomes, 14*, 1-7, doi:10.1186/s12955-016-0573-9.

Lorenzo-Seva, U., Ferrando, P. J., Calderon, C., del Mar Muñoz, M., Beato, C., Ghanem, I., et al. (2019). Psychometric properties and factorial analysis of invariance of the Satisfaction with Life Scale (SWLS) in cancer patients. *Quality of Life Research*, doi:10.1007/s11136-019-02106-y.

Mantel, N. (1963). Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Assosiation, 58*(303), 690-700.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal Of The National Cancer Institute, 22*(4), 719-748.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ, US: Lawrence Erlbaum.

Michalos, A., C. , & Kahlke, P. M. (2010). Stability and sensitivity in perceived quality of life measures: Some panel results. *Social Indicators Research, 98*, 403-434, doi:10.1007/s11205-009-9554-2.

Ortuño-Sierra, J., Aritio-Solana, R., Chocarro de Luis, E., Nalda, F. N., & Fonseca-Pedrero, E. (2019). Subjective well-being in adolescence: New psychometric evidences on the Satisfaction with Life Scale. *European Journal of Developmental Psychology, 16*(2), 236-244, doi:10.1080/17405629.2017.1360179.

Pavot, W., & Diener, E. (1993). Review of the Satisfaction With Life Scale. *Psychological assessment, 5*(2), 164-172, doi:10.1037/1040-3590.5.2.164.

Pavot, W., & Diener, E. (2008). The Satisfaction With Life Scale and the emerging construct of life satisfaction. *The Journal of Positive Psychology, 3*(2), 137-152, doi:10.1080/17439760701756946.

Petrillo, J., Cano, S. J., McLeod, L. D., & Coon, C. D. (2015). Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: A comparison of worked examples. *Value in Health, 18*(1), 25-34, doi:10.1016/j.jval.2014.10.005.

Proyer, R. T., Ruch, W., & Buschor, C. (2013). Testing strengths-based interventions: A preliminary study on the effectiveness of a program targeting curiosity, gratitude, hope, humor, and zest for enhancing life satisfaction. *Journal of Happiness Studies, 14*(1), 275-292, doi:10.1007/s10902-012-9331-9.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology, 5*, 27-48, doi:10.1146/annurev.clinpsy.032408.153553.

Rost, J., & Langeheine, R. (1997). A guide through latent structure models for categorical data. In J. Rost, & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 13-37). Münster, Germany: Waxman.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of personality and social psychology, 57*(6), 1069-1081, doi:10.1037/0022-3514.57.6.1069.

Schutte, L., Wissing, M. P., Ellis, S. M., Jose, P. E., & Vella-Brodrick, D. A. (2016). Rasch analysis of the Meaning in Life Questionnaire among adults from South Africa, Australia, and New Zealand. *Health And Quality Of Life Outcomes, 14*(1), 1-15, doi:10.1186/s12955-016-0414-x.

Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The Meaning in Life Questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology, 53*(1), 80-93, doi:10.1037/0022-0167.53.1.80.

Stephen, H., Lucy, T., & Paul, A. (2013). Measuring subjective well-being and its potential role in policy: Perspectives from the UK Office for National Statistics. *Social Indicators Research, 114*, 73-86, doi:10.1007/s11205-013-0384-x.

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis & Rheumatism: Arthritis Care & Research, 57*(8), 1358-1362, doi:10.1002/art.23108.

Vassar, M. (2008). A note on the score reliability for the Satisfaction with Life Scale: An RG study. *Social Indicators Research, 86*, 47-57, doi:10.1007/s11205-007-9113-7.

Vittersø, J., Biswas-Diener, R., & Diener, E. (2005). The divergent meanings of life satisfaction: Item response modeling of the Satisfaction with Life Scale in Greenland and Norway. *Social Indicators Research, 74*(2), 327-348, doi:10.1007/s11205-004-4644-7.

Wissing, M. P., & Van Eeden, C. (2002). Empirical clarification of the nature of psychological well-being. *South African Journal of Psychology, 32*(1), 32-44.

Table 1

Demographic profile

	South Africa	Italy
n	676	516
% female	64.05%	55.23%
Age: *M* (*SD*)	41.86 (11.86)	40.22 (11.59)
% tertiary education	63.17%	56.59%

Table 2

Rasch parameter estimates for the SWLS

	δ		SE		Infit		Outfit		*r*		Loading	
	SA	IT	SA	IT	SA	IT	SA	IT	SA	IT	SA	IT
1. In most ways my life is close to my ideal.	0.02	0.06	.04	.06	0.90	0.66	0.95	0.64	.75	.85	-.54	-.62
2. The conditions of my life are excellent.	-0.02	0.40	.04	.05	0.75	0.94	0.75	0.96	.78	.80	-.68	-.74
3. I am satisfied with my life.	-0.23	-0.47	.04	.06	0.69	0.64	0.68	0.63	.82	.87	-.49	-.26
4. So far I have gotten the important things I want in life.	-0.19	-0.40	.04	.06	1.07	1.12	1.06	1.11	.75	.80	.44	.45
5. If I could live my life over, I would change almost nothing.	0.42	0.40	.04	.05	1.48	1.58	1.49	1.54	.73	.80	.81	.76
Ideal values					*>0.6; <1.4*		*>0.6; <1.4*		>0			

Note. δ = average item challenge; SE = standard error of the item challenge; Infit = infit mean square statistic; Oufit = outfit mean square statistic; *r* = point-biserial correlation; Loading = Loading of the item residual on the first contrast of the Rasch PCA of the residuals; SA = South Africa; IT = Italy. This table is based on the original data without collapsing the response categories.

Table 3

Differential item functioning for country, gender and education level

Item	Country (Bonferroni α = .003)				Gender (Bonferroni α = .003)				Education level (Bonferroni α = .003)			
	DIF measure (DIF SE)		DIF contrast	MH test statistic (*p*-value)	DIF measure (DIF SE)		DIF contrast	MH test statistic (*p*-value)	DIF measure (DIF SE)		DIF contrast	MH test statistic (*p*-value)
	SA (*n*=676)	IT (*n*=516)			M (*n*=474)	F (*n*=718)			Secondary (*n*=473)	Tertiary (*n*=719)		
1	.04 (.05)	.04 (.05)	.00	1.527 (.217)	-.09 (.05)	.12 (.04)	-.21	12.944 (<.001)	.02 (.05)	-.01 (.04)	-.02	0.691 (.406)
2	-.03 (.05)	.33 (.05)	-.35	42.510 (<.001)	.09 (.05)	.17 (.04)	-.08	1.487 (.223)	.08 (.05)	-.03 (.04)	-.09	2.046 (.153)
3	-.26 (.05)	-.39 (.05)	.13	6.733 (.010)	-.32 (.06)	-.32 (.05)	.00	0.311 (.577)	-.36 (.06)	-.02 (.05)	-.07	1.499 (221)
4	-.22 (.05)	-.33 (.05)	.11	5.057 (.025)	-.14 (.05)	-.36 (.05)	.22	9.440 (.002)	-.22 (.06)	.02 (.05)	.08	1.772 (.183)
5	.48 (.04)	.33 (.05)	.15	4.717 (.030)	.47 (.05)	.37 (.04)	.10	1.685 (.194)	.47 (.05)	.03 (.04)	.10	1.721 (.190)

Note. Bonferroni α = Bonferroni corrected significance level; DIF = differential item functioning; DIF measure =item challenge for the particular group; DIF SE = standard error of item challenge for the particular group; MH = Mantel-Haenszel; SA = South Africa; IT = Italy; M = male; F = female. This table is based on the original data without collapsing the response categories.

Table 4

Differential item functioning for age group (Bonferroni α = .001)

Item	DIF measure (DIF SE)					Group comparison displaying minimum DIF	
	Group 1: 18-29 (*n* = 213)	Group 2: 30-39 (*n* = 339)	Group 3: 40-49 (*n* = 308)	Group 4: 50-60 (*n* = 270)	Group 5: > 60 (*n* = 62)	DIF contrast	MH test statistic (*p*-value)
1	.08 (.08)	-.03 (.06)	.01 (.07)	.14 (.07)	-.09 (.16)	2 vs 3: .04	2 vs 3: 0.001 (.983)
2	.08 (.08)	.14 (.06)	.14 (.07)	.19 (.07)	.01 (.15)	2 vs 3: .00	2 vs 3: 0.003 (.957)
3	-.39 (.08)	-.30 (.07)	-.26 (.07)	-.36 (.08)	-.29 (.16)	2 vs 5: .01	2 vs 4: 0.000 (.990)
4	-.08 (.08)	-.23 (.06)	-.27 (.07)	-.41 (.08)	-.58 (.17)	2 vs 3: .04	4 vs 5: 0.013 (.909)
5	-.32 (.08)	.41 (.06)	.39 (.06)	.41 (.07)	.84 (.14)	2 vs 4: .00	2 vs 3: 0.071 (.791)

Note. Bonferroni α = Bonferroni corrected significance level; DIF = differential item functioning; DIF measure = item challenge for the particular age group; DIF SE = standard error of item challenge for the particular age group; MH = Mantel-Haenszel. DIF contrast presented in absolute value. This table is based on the original data without collapsing the response categories.

1

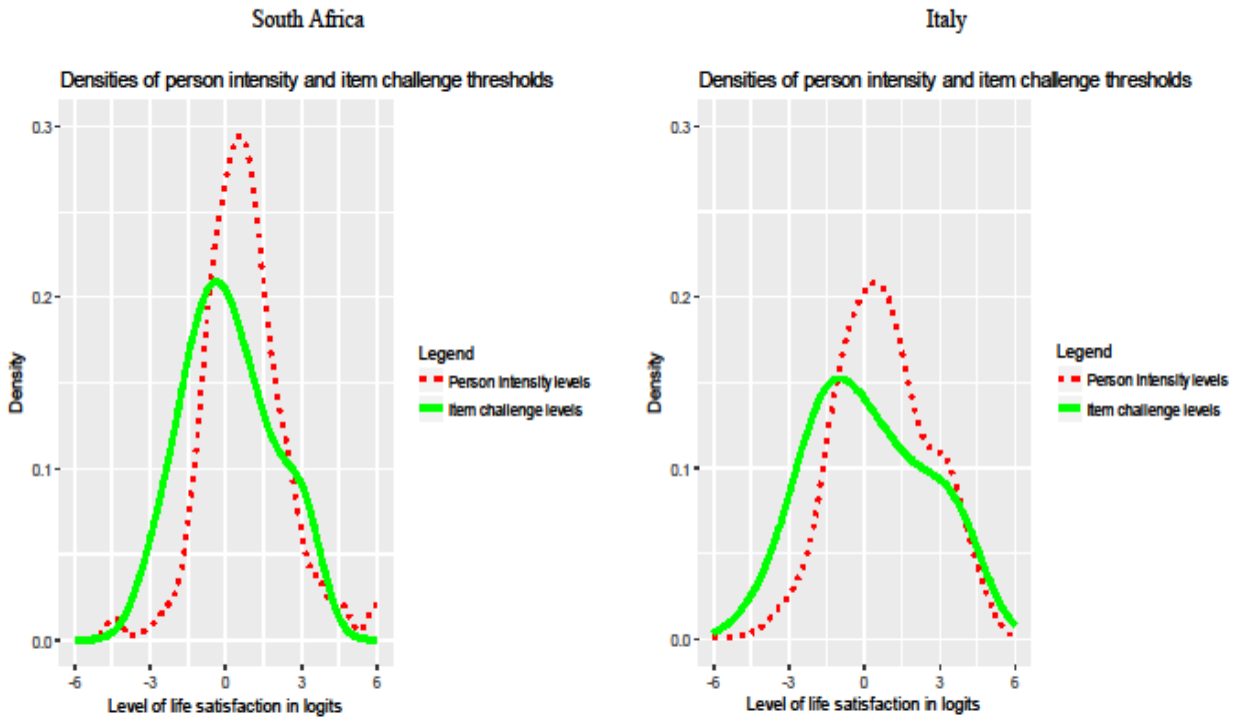


Figure 1. Density functions of the person intensity and item challenge threshold levels of the SWLS.

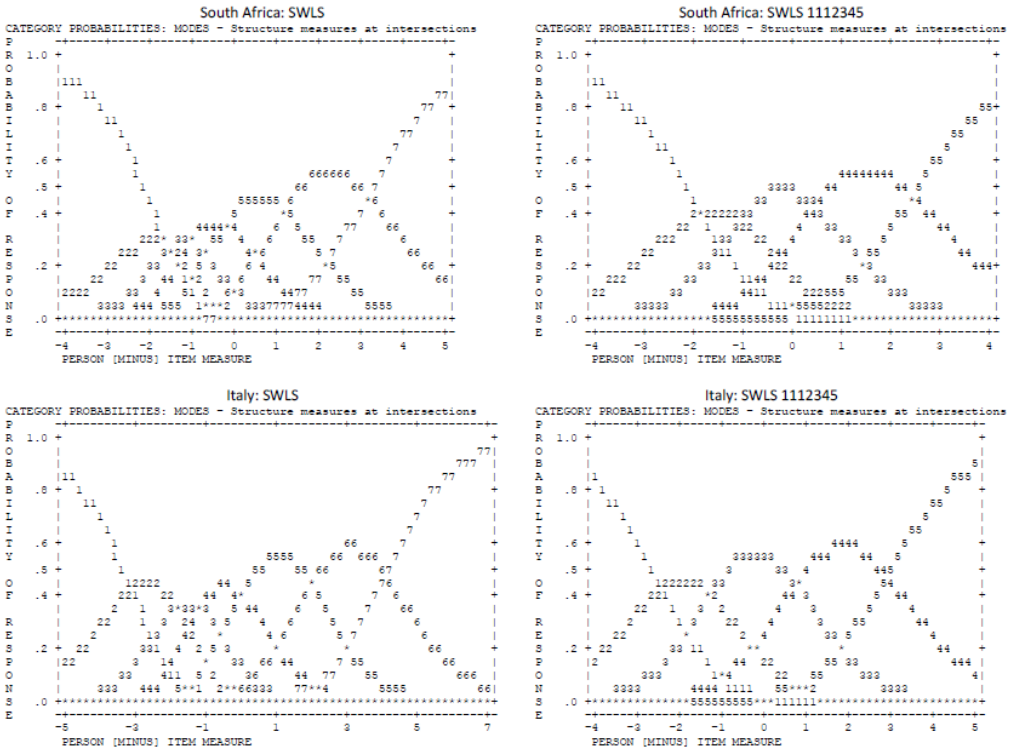


Figure 2. Category probability curves of the SWLS for South Africa and Italy. SWLS = Original SWLS; SWLS 1112345 = SWLS with response categories 1 to 3 collapsed.

ESM Table 1S

Rating Scale Category Functioning in South African and Italian Samples

	South Africa					Italy				
	Observed count	Average measure	Infit	Outfit	Threshold calibration	Observed count	Average measure	Infit	Outfit	Threshold calibration
1 – *Strongly disagree*	147	-1.38	1.75	2.32	None	74	-2.63	1.38	1.42	None
2 – *Disagree*	134	-1.28	0.78	0.82	-1.57	165	-1.78	1.03	1.1	-3.12
3 – *Slightly disagree*	270	-0.7	0.72	0.72	-1.51	279	-1.01	0.85	0.81	-1.86
4 – *Neither agree or disagree*	593	-0.09	0.76	0.73	-1.07	521	-0.12	0.88	0.88	-1.15
5 – *Slightly agree*	1006	0.68	0.78	0.79	-0.22	775	1	0.86	0.86	0.01
6 – *Agree*	869	1.66	0.85	0.86	1.21	557	2.45	1.07	1	1.99
7 – *Strongly agree*	361	2.57	1.61	1.26	3.17	209	3.56	1.26	1.15	4.12
Ideal values			*<2*	*<2*				*<2*	*<2*	

Note. Infit = infit mean square statistic; Oufit = outfit mean square statistic.

South Africa

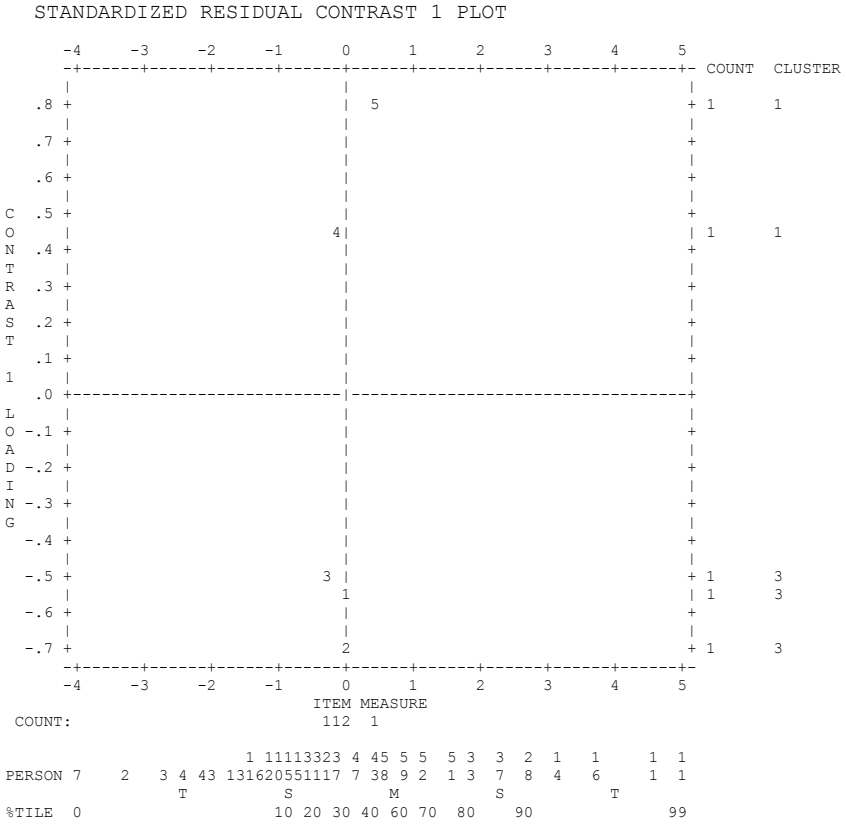
South Africa

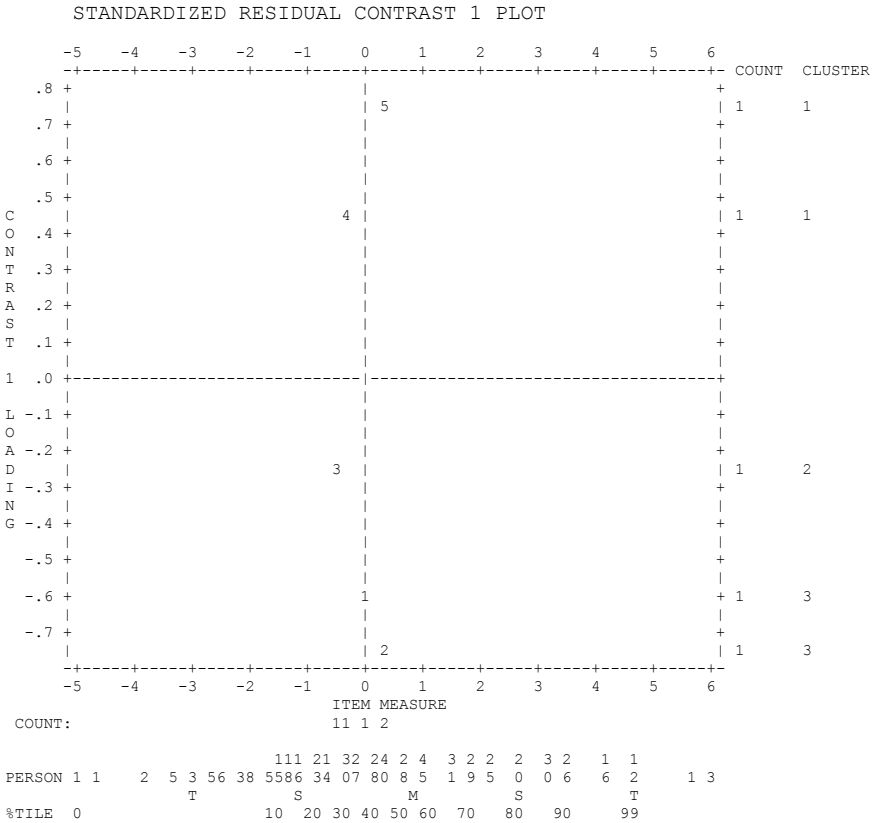
```
MEASURE    PERSON - MAP - ITEM
              <more>|<rare>
    5           .##  +
                     |
                .##  |
                     |
                     |
                     |
    4                +
               .###  |
                     |
                    T|
                     |
                .##  |
    3                +
                     |
             .#####  |
                     |
           .#######  |
                    S|
    2                +
            .######  |
                     |
        .##########  |
                     |
        .##########  |
    1                +
       .###########  |
       .########### M|
                     |T swls5
          .########  |
         .#########  |S
    0      .#######  +M swls1  swls2
                     |S swls3  swls4
              .####  |
            .######  |T
            .###### S|
             ######  |
   -1            ##  +
                .##  |
                 .#  |
                .##  |
                  .  |
                  .  |
   -2             . T+
                  .  |
                  .  |
                     |
                  .  |
                     |
   -3                +
                     |
                  .  |
                     |
                     |
                     |
   -4            .#  +
              <less>|<frequent>
 EACH "#" IS 5: EACH "." IS 1 TO 4
```

Italy

```
MEASURE    PERSON - MAP - ITEM
              <more>|<rare>
    6             .  +
                     |
                  .  |
                     |
                     |
    5                +
                ###  |
                     |
                    T|
                     |
    4          ####  +
                     |
            .######  |
                     |
           .#######  |
    3                +
                     |
              ##### S|
                     |
            .######  |
    2                +
           .#######  |
                     |
           .#######  |
                     |
    1  .###########  +
            ####### M|T
                     |
         ##########  |S swls2  swls5
            #######  |
    0                +M swls1
            .######  |
           .#######  |S swls3  swls4
               .###  |
             .#####  |T
   -1          #### S+
                     |
              .####  |
               .###  |
                 .#  |
   -2            ##  +
                  .  |
                 .#  |
                     |
                 .#  |
   -3             . T+
                     |
                 .#  |
                     |
                     |
   -4             .  +
                     |
                     |
                  .  |
                     |
   -5             .  +
              <less>|<frequent>
 EACH "#" IS 4: EACH "." IS 1 TO 3
```

ESM Figure 1S. Person-item maps of the Satisfaction with Life Scale for South Africa and Italy.



ESM Figure 2S. Standardised residual plot of the first contrast of the Rasch principal components analysis for South Africa and Italy.

Italy