Spanish Validation of the Flourishing Scale in the General Population

Irene Checa[1] · Jaime Perales[2] · Begoña Espejo[1]

Abstract Well-being research and its measurement have grown in the last two decades. The objective of this study was to adapt and validate the Flourishing Scale in a sample of Spanish adults. This was a cross-sectional study using a non-probabilistic sample of 999 Spanish general adult population participants. The psychometric properties of the scale were analysed from an exploratory and confirmatory perspective. Exploratory factor analysis showed a one-factor solution explaining 42.3% of the variance; an internal consistency of .846; temporal reliability correlation of .749; convergent validity with the Satisfaction with Life Scale of .521 and criterion validity with positive and negative affect (PANAS), pessimism and optimism (LOT-R) ranging from .270 to .488. Confirmatory factor analysis testing the one-factor solution showed a χ^2 of 65.57 df = 20; CFI of .982, RMSEA of .06, average variance extracted index of .518 and composite reliability index of .841. Results showed that the Spanish version of the FS is a reliable and valid method for measuring high levels of well-being.

Introduction

Well-being research and positive psychology have experienced a boom of popularity in the past couple of decades (Antolikova 2012; Helliwell et al. 2012). The debate between the ancient concept of eudemonic (Ryff and Keyes 1995; Ryan and Decy, 2001) and hedonic (Diener et al. 1985) well-being has led to the understanding of well-being as a dichotomy. Hedonic well-being refers to the pursuit of pleasure or happiness whereas eudemonic well-being refers more to human potential; the latter being a more global concept (Ryan et al. 2013).

While more interest has been paid to hedonic well-being in the last decade, measurements of well-being in the broader sense, especially measures of "flourishing" have recently gained attention (Hone et al. 2014a, b; Keyes 2002). Authors such as Felicia Huppert, Corey Keyes, Ed Diener and Martin Seligman have been key in developing this concept with their own peculiarities (Hone et al. 2014a, b). Keyes was the first to coin the term flourishing (Keyes 2002). Three of Keyes' domains of flourishing appear in every flourishing measure namely positive relationships, meaning or purpose and self-acceptance or self-esteem.

One of the most popular flourishing measures is Diener's Flourishing Scale (FS) (Diener et al. 2010). The original scale was comprised of 12 items although it was later reduced to eight. The 8-item FS showed sound psychometric properties in its development with an internal consistency of .87 and temporal reliabilities of .71. Factor analysis of the scale showed the scale had a unidimensional factor explaining

✉ Irene Checa
  irene.checa@uv.es

  Jaime Perales
  jperales@kumc.edu

  Begoña Espejo
  begonya.espejo@uv.es

[1] Departamento de Metodología de las Ciencias del Comportamiento, Universitat de València, Av. Blasco Ibáñes, 21, 46010 Valencia, Spain

[2] Department of Preventive Medicine and Public Health, University of Kansas Medical Center, 4125 Rainbow Blvd. MS 1056, Kansas City, KS 66103, USA

53% of the variance. The authors presented normative scores, item analyses and found correlations of .42 to .73 between the FS and the Basic Need Satisfaction in General Scale (Ryan and Deci 2000), Ryff's Scales of Psychological Well-being (Ryff and Keyes 1995) and the Satisfaction With Life Scale (Diener et al. 1985) indicating good convergent validity.

The FS has been validated in countries such as New Zealand (Hone et al. 2014a, b), Portugal (Silva and Caetano 2013), Japan (Sumi 2014a, b), China (Tang et al. 2014), France (Villieux et al. 2016) and Germany (Esch et al. 2013). All these validations have consistently found a one factor structure explaining between 44% (France) and 73.1% (Japan) of the variance. Convergent validity has been tested using instruments including the Subjective Happiness Scale (Lyubomirsky and Lepper 1999), Satisfaction with Life Scale (Diener et al. 1985), Revised Life Orientation Test (R-LOT) (Scheier et al. 1994), Positive and Negative Affect Schedule (Watson et al. 1988), Chinese Virtues Questionnaire (Duan et al. 2013) and Brief Symptom Inventory (Wang et al. 2013). Convergent validity correlations with the above mentioned scales ranged from .21 to .68 all of them being statistically significant. Two studies mention the testing of discriminant validity using the Hopkins Symptoms Checklist (Derogatis et al. 1974), Perceived Stress Scale (PSS) (Cohen et al. 1983) and The Centre for Epidemiological Studies Depression Scale (Radloff 1977). However, if these results were actually testing criterion validity as discriminant validity, it would mean that correlations among measures of the same concept measured with different methods need to be consistently higher than the correlations among measures of different concepts using the same method (Campbell and Fiske 1959). In addition, contrary to the results in the original version of the FS, all the other validations conduct confirmatory analyses showing good fit scores (CFI between .90 and .986, RMSEA between .041 and .08.). Regarding reliability, all validation studies show internal consistency scores between .83 and .95, although the French version does not report it.

Different well-being instruments have been validated and used in the Spanish context including the Satisfaction With Life Scale (SWLS) (Atienza et al. 2003), Day Reconstruction Method (DRM) (Caballero et al. 2014) and the Subjective Happiness Scale (SHS) (Extremera et al. 2011). All these instruments measure the individual perception of satisfaction using different formats of questions. A recent Spanish validation of the FS was conducted using non-probability samples from Spain and Colombia (Pozo-Muñoz et al. 2016). While this validation use samples from two Spanish-speaking countries and assess invariance, some limitations are to be noted. First, the authors did not carry out a back-translation of the items, as recommended by the International Test Commission Guidelines for test translation and adaptation (Muñiz et al. 2013; Muñiz et al. 2016). Second, Spanish and Colombian Spanish are different in some aspects, and the use of a translated scale into Spanish may not be equivalent in both countries. Third, although the authors assessed invariance between samples, cross-cultural adaptations require semantic, idiomatic, experiential and conceptual equivalence (Arafat et al. 2016; Muñiz et al. 2016), not only metric equivalence. Fourth, the Pozo-Muñoz version of the FS used a 5-point Likert scale rather than the 7-point Likert scale used by the original FS and all other validations (Diener et al. 2010; Hone et al. 2014a, b; Sumi 2014a; Silva and Caetano 2013; Tang et al. 2014) not justifying the change theoretically nor statistically. Other limitations include a fully academic sample, the lack of psychometric analysis of features such as convergent and criterion validity, temporal reliability and using the same sample to analyse exploratory and confirmatory factor structures rather than randomly splitting it in two as suggested in the literature (Brown 2015).

The present study aimed to address the limitations of the previous Spanish validation of the FS. Given the promising psychometric properties of the FS in different countries and populations, the objective of this paper was to validate the FS in a sample of Spanish adults. The psychometric properties of the scale were analysed from an exploratory and confirmatory perspective.

Methods

Design and Participants

The present project was a cross-sectional study using a non-probabilistic sample of the general population. Participants were recruited via email and social media including an explanation of the study and a link to LimeSurvey, an open source survey tool (LimeSurvey Project Team, Schmitz, Carsten 2012). Participants were included if they identified themselves as Spanish (irrespective of the country they lived in) and were at least 18 years of age. The final sample was comprised of 999 participants. Following standard guidelines for analysing the factorial structure of a scale, the sample was randomly split in two in order to conduct exploratory ($n = 502$) and confirmatory ($n = 497$) analyses (Anderson and Gerbing 1988; Brown 2015). The sample used for the retest was drawn from the total but the number of observations was reduced down to 102 individuals due to considerable attrition.

The average age of the total sample was 28.4 (SD = 11.7) ranging from 18 to 71 years of age. Of the sample, 31.3% were men, 96.0% had completed at least college/university education and 34.1% were married or cohabiting. Regarding the working status, 41.9% were students only, 27.3% were employed only, 24.9% were employed and students at the same time, 3.6% were unemployed, 1% inactive and 3.6% retired. Regarding Sample 1, the average age was 28.98 (SD = 11.99) and 31.3% were men. The mean age of

Sample 2 was 27.77 (SD = 11.27) and 31.7% were men. There were no statistical differences between socio-demographic characteristics of the two samples (see Table 1).

Measures

The Flourishing Scale (Diener et al. 2010) is an eight-item instrument describing important aspects of human functioning including positive relationships, feelings of competence and having meaning and purpose in life. The instrument uses a seven-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). Total scores range from 8 to 56 with high scores indicating respondents viewing themselves in positive terms in important areas of functioning.

Convergent Validity

The Satisfaction With Life Scale (SWLS) is a 5-item instrument designed to measure global cognitive judgment of satisfaction with one's life (Diener et al. 1985). Participants indicate how much they agree or disagree with each of the 5 items using a seven-point scale that ranges from seven (strongly agree) to one (strongly disagree). Items include 'In most ways my life is close to my ideal' or 'I am satisfied with my life'. Scores range from 7 to 35, higher scores indicating higher satisfaction. The Spanish version used in the present study has been validated in a general sample of Spanish adults and shown an internal consistency of .858 (Atienza et al. 2003). In our total sample, internal consistency was .858.

Criterion Validity

The Life Orientation Test-Revised (LOT-R) has been used to measure optimism (Scheier et al. 1994). The scale is comprised of ten items, four measuring control, three pessimism and three optimism. Each item of the LOT-R is answered on a 1–5 scale that ranges from strong disagreement to strong agreement. Examples of items from this scale are 'In uncertain times, I usually expect the best' or 'I rarely count on good things happening to me'. The scores range from 0 to 12, higher scores indicating higher optimism and lower pessimism. The Spanish version used in the present study has been validated in a general sample of Spanish adults and has shown sound psychometric properties (Ferrando et al. 2002). In our total sample, internal consistency was .763 for optimism and .701 for pessimism.

The Positive And Negative Affect Schedule (PANAS) measures two domains, positive and negative affect (Watson et al. 1988). This 20-item instrument uses a five point Likert scale ranging from 1 (not at all) to 5 (extremely). Items include 'interested', 'excited', 'ashamed' or 'irritable'. Scores range from 10 to 50, higher scores indicating higher affect. The Spanish version used in the present study has been validated in a large sample of Spanish adolescents and young adults showing internal consistency scores from .80 to .86 (Ortuño-Sierra et al. 2015). In our total sample, internal consistency was .845 for positive affect and .855 for negative affect.

Procedure

The Spanish adaptation process of the FS was conducted using criteria proposed by Muñiz et al. (2013, 2016). The authors stress the need to ensure conceptual, linguistic and metric equivalence. The adaptation of the scale was conducted using the translation-back translation method by two bilingual translators. Discrepancies between both translators were reported to the original author, Ed Diener, for clarification. Later, two qualified judges studied the conceptual and linguistic meaning of the item in Spanish for refinement. The final Spanish version of the FS is shown in Appendix 1. Participants who accepted to enrol in the study had to read and accept an online inform consent before entering the study. Data collection complied with the Spanish Law of Data ensuring confidentiality and anonymity. Data collection started in February 2015 finishing in September 2015. A retest was conducted one month later in order to obtain temporal reliability data. The study met the standards set by the *Comité de Ética de la Investigación y Docencia* at the *Universidad de Valencia (UV)*.

Data Analysis

A first sample was used to conduct an exploratory factor analysis (EFA) and calculate reliability and item analysis of the FS. Temporal reliability was obtained by calculating the correlation between the total FS score at baseline and the retest total FS score one month later. Convergent validity was measured by calculating the correlation between the total FS score and the SWLS. It was hypothesised that high scores of the FS would be associated with high SWLS scores. Criterion validity was measured using simple linear regression including the total FS score along with scores of optimism and pessimism (LOT-R) and positive affect and negative affect (PANAS). It was hypothesised that higher FS scores would be associated with higher scores in optimism and positive affect and lower scores in pessimism and negative affect.

A second sample was used to conduct confirmatory factor analysis (CFA) testing the one-dimensional model as suggested in the literature (Diener et al. 2010). Polychoric correlations and robust χ^2 fitting indicators were calculated given the number of response options and the non-parametric nature of the data (Babakus et al. 1987; DiStefano 2002; Finney and DiStefano 2006; Hutchinson and Olmos 1998; Jöreskog et al. 1999). Due to the sensitivity of the sample size of the χ^2 goodness of fit test, we used the Comparative Fit Index (CFIs) and Root Mean Square Error of Approximation

Table 1 Socio-demographic characteristics of the total sample and the two subsamples

		Total Sample (n = 999)	Sample 1 (n = 502)	Sample 2 (n = 497)	Test statistic
Age	Mean (SD)	28.4 (11.7)	28.98 (11.99)	27.77 (11.27)	1.645
Gender	Men	31.3%	31.7%	31.0%	0.24
	Women	68.7%	68.3%	69.0%	-0.24
Education	None or primary	1.9%	2.4%	1.4%	1.16
	Secondary	2.1%	2.4%	1.8%	0.66
	GED	55.3%	55.2%	55.4%	-0.06
	College-Master's	28.5%	28.5%	28.6%	-0.03
	Doctorate	12.1%	11.6%	12.7%	-0.53
Marital status	Single	63.0%	63.1%	62.9%	0.03
	Married or cohabiting	34.1%	33.3%	34.9%	-0.53
	Divorced	2.6%	3.2%	2.0%	1.19
	Widow	0.3%	0.4%	0.2%	1.85
Working status	Student only	41.9%	40.4%	43.3%	-0.93
	Student and worker	24.9%	24.7%	25.2%	-0.18
	Paid-employed	22.8%	23.1%	22.6%	0.19
	Self-employed	4.5%	5.2%	3.8%	1.07
	Unemployed	4.6%	5.2%	4.0%	0.90
	Retired	1.2%	1.4%	1.0%	0.58

No test comparing Sample 1 vs Sample 2 was found statistically significant at the .05 level

(RMSEA) to determine model fit. CFI values of .90 or above and RMSEA values above .06 and below .08 are indicative of good empirical fit (Schumacker and Lomax 2004). Reliability in this confirmatory sample was calculated using the Composite Reliability Index and the Average Variance Extracted Index (Raykov 1997). Data analyses were conducted using SPSS 22.0 (SPSS, I. 2013) and EQS 6 (Bentler 2006).

Results

Results from the exploratory factor analysis (EFA) can be found in Table 2. Using principal axis factoring with the eight FS items, we found good sample adequacy (KMO = .876; χ^2 de Barlett = 1377.923; p = .000). EFA led to a one-dimensional factor solution explaining 42.3% of the variance and all saturations ranged from .540 to .730. Internal consistency alpha was .846 not improving with the deletion of any item. Corrected item-total correlations ranged from .356 to .493. Inter-item correlation also seemed adequate, ranging from a minimun of .318 to a maximum of .570.

Temporal reliability was calculated using 102 participants who completed the retest survey. Test-retest correlation was .749, p < .001. Regarding convergent validity, correlation with SWLS was .521, p < .001. Regarding criterion validity, correlations with positive affect, optimism, negative affect and pessimism were .422, .488, −.270 and −.385 respectively. All of them were all significant and in the expected direction.

Confirmatory factor analysis (CFA) was carried out to test the one-dimensional model. CFA's χ^2 was significant (χ^2 = 65.5765; df = 20) although the goodness of fit indicators were satisfactory (CFI = .982 y RMSEA = .06). All saturations ranged from .53 to .74. The model's goodness of fit was tested again after deleting item 1, which showed a lower variability and saturation than the rest of the items. The deletion of item 1 did not improve the model (CFI = .981; RMSA = .08). The Average Variance Extracted Index was .518 and the Composite Reliability Index was .841.

Discussion

The aim of this study was to adapt the FS to Spanish and to validate it in a sample of Spanish adults. The psychometric properties have been analysed using standard criteria (Muñiz et al. (2013, 2016). Overall, the present Spanish-language adaptation of the FS showed adequate reliability and criterion validity. Both exploratory and confirmatory factor analyses provided support for a one-dimensional flourishing construct using an eight-item scale. This one-factor solution has been found previously in other studies using the same scale. The present adaptation has addressed many of the limitations of a recent Spanish translation validated in Colombia and Spain (Pozo-Muñoz et al. 2016). The limitations addressed include conducting the back-translation of the scale, using a general population sample, analysing convergent and criterion

Table 2 Item analysis of the FS in sample 1

| Item | Mean | SD | HI | α if item deleted | Item correlation | | | | | | |
					2	3	4	5	6	7	8
1	6.50	.873	.463	.841	.318	.350	.320	.370	.453	.232	.336
2	5.66	1.250	.383	.827		.463	.414	.339	.517	.376	.482
3	5.87	1.086	.493	.821			.570	.564	.488	.336	.417
4	5.82	.944	.394	.829				.431	.467	.307	.398
5	6.18	.973	.430	.825					.449	.446	.431
6	5.91	1.051	.461	.817						.412	.486
7	5.38	1.444	.356	.840							.527
8	5.69	1.132	.429	.820							
TOTAL				α = .846							

n = 502; *SD* Standard Deviation, *HI* Homogeneity Index

validity, temporal reliability and randomly splitting it in two to analyse exploratory and confirmatory factor structures.

Our study showed high scores in FS items which has also been found in previous studies including that of the original scale (Diener et al. 2010). The FS was created to measure high levels of well-being (Keyes 2002). Therefore, in a continuum of psychological well-being, flourishing would be located on the most positive side. Only the Japanese version found slightly lower scores. These lower scores could be related to a different cultural meaning of well-being (Kitayama et al. 2000).

The exploratory factor analysis in this study has several differences and similarities with previous validations of the FS scale. This study found a lower percentage of explained variance compared to previous studies (Esch et al. 2013; Hone et al. 2014a, b; Tang et al. 2014) and similar to the French version (Villieux et al. 2016). Item 1 showed a low variability and a high mean compared to the other items on the scale. Low variability and a high mean may compromise the scale's cumulative error and could be one of the reasons why the percentage of explained variance is not as high as previous versions of the scale even though it showed good evidence of construct validity. Regarding other types of validity, the present study found similar levels of convergent and criterion validity as well as similar internal consistency compared with previous validations (Silva and Caetano 2013; Sumi 2014a). Temporal reliability is higher than the original validation of the scale (Diener et al. 1985) but lower than japanese study (Sumi 2014b).

Regarding the confirmatory factor analysis, the Spanish sample showed adequate goodness of fit, and the Average Variance Extracted Index and the Composite Reliability Index calculated were satisfactory. The CFI had values that were higher than those found in the Portugal, New Zealand and China validations (Hone et al. 2014a, b; Silva and Caetano 2013; Tang et al. 2014), but lower compared the Japanese validation (Sumi 2014a). The residual RMSEA values were lower than the ones found in the Japan and New Zealand validations (Hone et al. 2014a, b; Sumi 2014a). In the latter adaptation, the RMSEA was found not to be adequate and only showed an acceptable value that was correlated with some items' errors (Hone et al. 2014a, b).

This study has two main implications. First, the FS can now be used to assess efficacy of interventions aiming to increase well-being as a primary or secondary aim. Trials have shown that interventions such as a six-week yoga intervention (Manincor et al. 2016), a web-based happiness training (Feicht et al. 2013) and body-mind medicine (Gimpel et al. 2014) can increase FS scores in different populations and settings. Second, there has been an increasing interest from governments to measure well-being at a national and international level in the past decade (Helliwell 2003). The FS scale is being adapted and validated in different countries and languages. Therefore, the FS scale has the potential to be used in wide international epidemiological studies to compare levels of well-being and to assess risk and protective factors. The use of this scale would suppose a quick, novel and psychometrically sound way to measure well-being.

This study has some limitations. First, the paucity of retest data made it impossible to assess confirmatory temporal reliability. Second, the present study did not assess discriminant validity. Further evidence is needed to test whether several constructs that are supposed to be unrelated to flourishing such as affect intensity and impulsivity, are indeed unrelated (Diener et al. 1985). Third, the sample used in this study was a non-probabilistic one. In order to prove external validity, the present scale should be validated in a representative sample of the Spanish population.

Further studies could assess the suitability of the response scale as the literature suggests seven-point Likert scales are excessive and not more efficient compared to five (Hartley 2014; Lai et al. 2010; McDonald 2004). Additionally, many studies question central values in response scales given that the verbal statement that acompanies the central number does not mean the individual considers they are in the middle of the continuum of the corresponding latent construct (Dalal et al. 2014; González-Romá and Espejo 2003; Hernández et al.

2006; Kulas et al. 2008; Kulas and Stachowski 2009; Murray et al. 2015; Onwuegbuzie and Weems 2004). Also, the probability of chosing a central value in the scale is much lower than chosing any other value even for people whose trait levels are central (González-Romá and Espejo 2003; Hernández et al. 2006). If future studies showed that reducing the number of response options were possible, the scale would be more parsimonious and would increase the validity of the responses without decreasing reliability. Future studies could also analyse the item characteristics from the Item Response Theory perspective. These studies would help analyse the information contributed by each item and decide whether item 1 provides enough information to keep it in the scale. Also, psychometric equivalence studies could be conducted using different versions of the scale in order to test whether there is a differential functioning item.

Conclusion

There is enough evidence to conclude that the Spanish version of the FS is a reliable and valid method for measuring high levels of well-being. The FS is a scale that is short and easy to understand and can be used for different research designs including trials and large surveys. This study has set the grounds for these different study designs in Spain and for clinical use.

Compliance with Ethical Standards

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Conflict of Interest The authors declare that they have no conflict of interest.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Appendix 1. Spanish adaptation of the FS

A continuación, encontrará ocho afirmaciones con las que usted puede estar de acuerdo o no. Usando la escala de 1 (Totalmente en desacuerdo) a 7 (Totalmente de acuerdo) que se le proporciona, indique en qué grado de acuerdo o desacuerdo está para cada una de las afirmaciones.

1. Mi objetivo es conseguir una vida plena y significativa/ I lead a purposeful and meaningful life

2. Mis relaciones sociales son gratificantes y me ofrecen el apoyo que necessito/ My social relationships are supportive and rewarding

3. Me implico y me intereso en mis actividades diarias/ I am engaged and interested in my daily activities

4. Contribuyo activamente a la felicidad y el bienestar de otros/ I actively contribute to the happiness and well-being of others

5. Soy competente y capaz en las tareas que son importantes para mí/ I am competent and capable in the activities that are important to me

6. Soy una buena persona y tengo una buena vida/ I am a good person and live a good life

7. Veo mi futuro con optimismo/ I am optimistic about my future

8. La gente me respeta/ People respect me

References

Anderson, J. C., y Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*(3), 411–423. doi:10.1037/0033-2909.103.3.411

Antolikova, E. (2012). The European year for active ageing and solidarity between generations 2012. *Kniznica, 13*(2), 6–9 Retrieved from http://goo.gl/M7EYfa.

Arafat, S. Y., Chowdhury, H. R., Qusar, M. S., & Hafez, M. A. (2016). Cross Cultural Adaptation & Psychometric Validation of research instruments: A methodological review. *Journal of Behavioral Health, 5*, 129–136. doi:10.5455/jbh.20160615121755.

Atienza, F. L., Balaguer, I., & García-Merita, M. L. (2003). Satisfaction with life scale: Analysis of factorial invariance across sexes. *Personality and Individual Differences, 35*(6), 1255–1260. doi:10.1016/S0191-8869(02)00332-X.

Babakus, E., Ferguson Jr., C. E., & Jöreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research, 24*(2), 222–228. doi:10.2307/3151512.

Bentler, P. M. (2006). *EQS: Structural equations program manual*. Encino: Multivariate Software, Inc. Retrieved from http://goo.gl/szIGcN.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Caballero, F. F., Miret, M., Olaya, B., Perales, J., López-Ridaura, R., Haro, J. M., Chatterji, S., & Ayuso-Mateos, J. L. (2014). Evaluation of affect in mexico and spain: Psychometric properties and usefulness of an abbreviated version of the day reconstruction method. *Journal of Happiness Studies, 15*(4), 915–935. doi:10.1007/s10902-013-9456-5.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81–105. doi:10.1037/h0046016.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*(4), 385–396 Retrieved from http://goo.gl/ftIgSk.

Dalal, D. K., Carter, N. T., & Lake, C. J. (2014). Middle response scale options are inappropriate for ideal point scales. *Journal of Business and Psychology, 29*(3), 463–478. doi:10.1007/s10869-013-9326-5.

Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhuth, E. H., & Covi, L. (1974). The Hopkins symptom checklist (HSCL): A self report symptom inventory. *Behavioral Science, 19*(1), 1–15. doi:10.1002/bs.3830190102.

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., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research, 97*(2), 143–156. doi:10.1007/s11205-009-9493-y.

DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling, 9*(3), 327–346.

Duan, W., Ho, S. M. Y., Bai, Y., & Tang, X. (2013). Psychometric evaluation of the Chinese virtues questionnaire. *Research on Social Work Practice, 23*(3), 336–345. doi:10.1177/1049731513477214.

Esch, T., Jose, G., Gimpel, C., Von Scheidt, C., & Michalsen, A. (2013). Die Flourishing Scale (FS) von Diener et al. liegt jetzt in einer autorisierten deutschen Fassung (FS-D) vor: Einsatz bei einer Mind-Body-medizinischen Fragestellung. *Forschende Komplementarmedizin, 20*(4), 267–275. doi:10.1159/000354414.

Extremera, N., Salguero, J. M., & Fernández-Berrocal, P. (2011). Trait meta-mood and subjective happiness: A 7-week prospective study. *Journal of Happiness Studies, 12*(3), 509–517. doi:10.1007/s10902-010-9233-7.

Feicht, T., Wittmann, M., Jose, G., Mock, A., von Hirschhausen, E., & Esch, T. (2013). Evaluation of a seven-week web-based happiness training to improve psychological well-being, reduce stress, and enhance mindfulness and flourishing: A randomized controlled occupational health study. *Evidence-based Complementary and Alternative Medicine, 2013*, 1–14. doi:10.1155/2013/676953.

Ferrando, P. J., Chico, E., & Tous, J. M. (2002). Propiedades psicométricas del test de optimismo Life Orientation Test. *Psicothema, 14*(3), 673–680 Retrieved from http://goo.gl/OMlGtS.

Finney, A.J. y DiStefano C. (2006). Non-normal and categorical data in structural equation modeling. In: G. R. Hancock y R. O. Mueller (Eds). Structural equation modeling: A second course, 269–314. Greenwich: Information Age Publishing

Gimpel, C., von Scheidt, C., Jose, G., Sonntag, U., Stefano, G. B., Michalsen, A., & Esch, T. (2014). Changes and interactions of flourishing, mindfulness, sense of coherence, and quality of life in patients of a mind-body medicine outpatient clinic. *Forschende Komplementarmedizin, 21*(3), 154–162. doi:10.1159/000363784.

González-Romá, V., & Espejo, B. (2003). Testing the middle response categories" not sure"," in between" and"?" in polytomous items. *Psicothema, 15*(2), 278–284.

Hartley, J. (2014). Some thoughts on likert-type scales. *International Journal of Clinical and Health Psychology, 14*(1), 83–86. doi:10.1016/S1697-2600(14)70040-7.

Helliwell, J. F. (2003). How's life? Combining individual and national variables to explain subjective well-being. *Economic Modelling, 20*(2), 331–360.

Helliwell, J., Layard, R., & Sachs, J. (2012). World happiness report [2012]. doi:10.14288/1.0053622

Hernández, A., Espejo, B., & González-Romá, V. (2006). The functioning of central categories middle level and sometimes in graded response scales: Does the label matter? *Psicothema, 18*(2), 300–306.

Hone, L., Jarden, A., & Schofield, G. (2014b). Psychometric properties of the flourishing scale in a new zealand sample. *Social Indicators Research, 119*(2), 1031–1045. doi:10.1007/s11205-013-0501-x.

Hone, L. C., Jarden, A., Schofield, G., & Duncan, S. (2014a). Measuring flourishing: The impact of operational definitions on the prevalence of high levels of wellbeing. *International Journal of Wellbeing, 4*(1), 62–90. doi:10.5502/ijw.v4i1.4.

Hutchinson, S. R., & Olmos, A. (1998). Behavior of descriptive fit indexes in confirmatory factor analysis using ordered categorical data. *Structural Equation Modeling: A Multidisciplinary Journal, 5*(4), 344–364. doi:10.1080/10705519809540111.

Jöreskog, K. G., Sörbom, D., Du Toit, S., & Du Toit, M. (1999). LISREL 8: New statistical features, chicago, IL: Scientific software international. *Inc.[Links],*

Keyes, C. L. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of Health and Social Behavior, 43*(2), 207–222.

Kitayama, S., Markus, H. R., & Kurokawa, M. (2000). Culture, emotion, and well-being: Good feelings in Japan and the United States. *Cognition & Emotion, 14*(1), 93–124. doi:10.1080/026999300379003.

Kulas, J. T., & Stachowski, A. A. (2009). Middle category endorsement in odd-numbered likert response scales: Associated item characteristics, cognitive demands, and preferred meanings. *Journal of Research in Personality, 43*(3), 489–493.

Kulas, J. T., Stachowski, A. A., & Haynes, B. A. (2008). Middle response functioning in likert-responses to personality items. *Journal of Business and Psychology, 22*(3), 251–259.

Lai, M., Li, Y., & Liu, Y. (2010). Determining the optimal scale width for a rating scale using an integrated discrimination function. *Measurement, 43*(10), 1458–1471.

LimeSurvey Project Team, Schmitz, Carsten. (2012). *LimeSurvey: An open source survey tool.* Germany: LimeSurvey Project Hamburg.

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research, 46*(2), 137–155. doi:10.1023/A:1006824100041.

Manincor, M., Bensoussan, A., Smith, C. A., Barr, K., Schweickle, M., Donoghoe, L., Bourchier, S., & Fahey, P. (2016). Individualized yoga for reducing depression and anxiety, and improving well-being: A randomized controlled trial. *Depression and Anxiety.* doi:10.1002/da.22502.

McDonald, J. L. (2004). The optimal number of categories for numerical rating scales. *Dissertation Abstracts International, 65*, 1664.

Muñiz, J., Elosúa, P., & Hambleton, R. K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema, 25*(2), 151–157. doi:10.7334/psicothema2013.24.

Muñiz, J., Elosua, P., Padilla, J., & Hambleton, R. K. (2016). Test adaptation standards for cross-lingual assessment. In C. S. Wells, & M. Faulkner-Bond (Eds.), *Educational measurement: From foundations to future; educational measurement: From foundations to future* (pp. 291-304, Chapter xvii, 494 Pages) Guilford Press, New York. Retrieved from http://search.proquest.com/docview/1828875838?accountid=14777

Murray, A. L., Booth, T., & Molenaar, D. (2015). When middle really means "top" or "bottom": An analysis of the 16PF5 using bock's nominal response model. *Journal of Personality Assessment*, 1–13. doi:10.1080/00223891.2015.1095197.

Onwuegbuzie, A., & Weems, G. (2004). Response categories on rating scales: Characteristics of item respondents who frequently utilize midpoint. *Research in the Schools, 11*(1), 50–59.

Ortuño-Sierra, J., Santarén-Rosell, M., Albéniz, A. P. d., & Fonseca-Pedrero, E. (2015). Dimensional structure of the spanish version of the positive and negative affect schedule (PANAS) in adolescents and young adults. *Psychological Assessment, 27*(3), e1–e9. doi:10.1037/pas0000107.

Pozo-Muñoz, C., Garzón-Umerenkova, A., Bretones-nieto, B., & Ligia-Charry, C. (2016). Psychometric properties and dimensionality of the "flourishing scale" in Spanish-speaking population. *Electronic Journal of Research in Educational Psychology, 14*(1), 175–192.

Radloff, L. S. (1977). The CES-D scale a self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*(3), 385–401. doi:10.1177/014662167700100306.

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement, 21*(2), 173–184. doi:10.1177/01466216970212006.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. doi:10.1037/0003-066X.55.1.68.

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology, 52*(1), 141–166. doi:10.1146/annurev.psych.52.1.141.

Ryan, R. M., Huta, V., & Deci, E. L. (2013). Living well: A self-determination theory perspective on eudaimonia. *The exploration of happiness* (pp. 117-139) Springer Netherlands.

Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology, 69*(4), 719. doi:10.1037/0022-3514.69.4.719.

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the life orientation test. *Journal of Personality and Social Psychology, 67*(6), 1063–1078. doi:10.1037/0022-3514.67.6.1063.

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. Lawrence Erlbaum Associates: New Jersey.

Silva, A. J., & Caetano, A. (2013). Validation of the flourishing scale and scale of positive and negative experience in Portugal. *Social Indicators Research, 110*(2), 469–478. doi:10.1007/s11205-011-9938-y.

SPSS, I. (2013). *IBM SPSS statistics 22*. New York: IBM Corp.

Sumi, K. (2014a). Reliability and validity of japanese versions of the flourishing scale and the scale of positive and negative experience. *Social Indicators Research, 118*(2), 601–615. doi:10.1007/s11205-013-0432-6.

Sumi, K. (2014b). Temporal stability of the japanese versions of the flourishing scale and the scale of positive and negative experience. *Journal of Psychology & Psychotherapy, 4*, 140. doi:10.4172/2161-0487.1000140.

Tang, X., Duan, W., Wang, Z., & Liu, T. (2014). Psychometric evaluation of the simplified chinese version of flourishing scale. *Research on Social Work Practice, 26*(5), 591–599. 5. doi:10.1177/1049731514557832.

Villieux, A., Sovet, L., Jung, S. C., & Guilbert, L. (2016). Psychological flourishing: Validation of the French version of the flourishing scale and exploration of its relationships with personality traits. *Personality and Individual Differences, 88*, 1–5. doi:10.1016/j.paid.2015.08.027.

Wang, J., Kelly, B. C., Liu, T., Zhang, G., & Hao, W. (2013). Factorial structure of the brief symptom inventory (BSI)-18 among chinese drug users. *Drug and Alcohol Dependence, 133*(2), 368–375. doi:10.1016/j.drugalcdep.2013.06.017.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. doi:10.1037/0022-3514.54.6.1063.