Article

Measurement Invariance of the Scale of Positive and Negative Experience Across 13 Countries

Assessment
2022, Vol. 29(7) 1507–1521

DOI: 10.1177/10731911211021494
journals.sagepub.com/home/asm
SAGE

Veljko Jovanović[1], Mohsen Joshanloo[2], Marta Martín-Carbonell[3], Corrado Caudek[4], Begoña Espejo[5], Irene Checa[5], Julia Krasko[6], Theodoros Kyriazos[7], Jarosław Piotrowski[8], Sean P. M. Rice[9], Ana Junça Silva[10,11], Kamlesh Singh[12], Katsunori Sumi[13], Kwok Kit Tong[14], Murat Yıldırım[15], and Magdalena Żemojtel-Piotrowska[8]

Abstract

The Scale of Positive and Negative Experience (SPANE) is widely used to measure emotional experiences, but not much is known about its cross-cultural utility. The present study evaluated the measurement invariance of the SPANE across adult samples (N = 12,635; age range = 18-85 years; 58.2% female) from 13 countries (China, Colombia, Germany, Greece, India, Italy, Japan, Poland, Portugal, Serbia, Spain, Turkey, and the United States). Configural and partial scalar invariance of the SPANE were supported. Three items capturing specific negative emotions (sad, afraid, and angry) were found to be culturally noninvariant. Our findings suggest that the SPANE's positive emotion terms and general negative emotion terms (e.g., negative and unpleasant) might be more suitable for cross-cultural studies on emotions and well-being, whereas caution is needed when comparing countries using the SPANE's specific negative emotion items.

Keywords

SPANE, positive emotions, negative emotions, measurement invariance, multigroup confirmatory factor analysis, cross-cultural

The tripartite model of subjective well-being (SWB; Diener, 1984; Diener et al., 1999) is one of the most extensively studied models in the field of well-being (Diener et al., 2018). According to this model, SWB includes three distinct components: life satisfaction, positive affect, and negative affect. To date, a number of scales have been developed to measure SWB (Cooke et al., 2016; Linton et al., 2016). The most frequently used scale for the assessment of life satisfaction is the Satisfaction with Life Scale (Diener et al., 1985), whereas two of the most widely used scales for the assessment of affective component of SWB are the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988), and the more recently developed Scale of Positive and Negative Experience (SPANE; Diener et al., 2010). Although the vast majority of SWB studies focus on samples drawn from a single country, cross-national and cross-cultural studies on levels and determinants of well-being, as well as on the performance of SWB scales across different cultural contexts have gained increasing popularity over the past few decades. The cross-cultural measurement invariance of the Satisfaction with Life Scale has been extensively evaluated in cross-national studies (Jang et al., 2017; Jovanović & Brdar, 2018). A cross-culturally validated version of the PANAS has also been developed, the International Positive and Negative Affect Schedule Short-form (Thompson, 2007). In contrast, the cross-cultural consistency of the SPANE has not been previously investigated. Therefore, the present study aims to fill this gap in the

[1]University of Novi Sad, Novi Sad, Serbia
[2]Keimyung University, Daegu, South Korea
[3]Universidad Cooperativa de Colombia, Santa Marta, Colombia
[4]Università degli Studi di Firenze, Florence, Italy
[5]University of Valencia, Valencia, Spain
[6]Ruhr University Bochum, Bochum, Germany
[7]Panteion University, Athens, Greece
[8]Cardinal Stefan Wyszynski University, Warsaw, Poland
[9]Oregon Health & Science University, Portland, OR, USA
[10]Instituto Universitário de Lisboa, Lisbon, Portugal
[11]Instituto Politécnico de Tomar, Tomar, Portugal
[12]Indian Institute of Technology Delhi, New Delhi, India
[13]Nagoya Institute of Technology, Nagoya, Japan
[14]University of Macau, Macau
[15]Ağrı İbrahim Çeçen University, Ağrı, Turkey

Corresponding Author:
Veljko Jovanović, Department of Psychology, Faculty of Philosophy, University of Novi Sad, Dr Zorana Đinđića 2, Novi Sad 21 000, Serbia.
Email: veljko.jovanovic@ff.uns.ac.rs

literature by examining the measurement invariance of the SPANE across different cultural contexts. As argued by Cieciuch et al. (2019):

> Before any cross-national comparisons are conducted, it is necessary to ensure that the same latent variables are measured in different countries, that respondents understand the items in a similar manner, and that they use the response scales in the same way. (p. 159)

In other words, to allow for a meaningful comparison in cross-national research, a scale needs to show evidence of measurement invariance (i.e., it needs to measure the same construct in the same way across groups, and its items need to have the same, or a highly similar, meaning across different groups). If the scale items are noninvariant, then the results of cross-country comparisons might be compromised, inaccurate, and misleading (Putnick & Bornstein, 2016).

Overview of the SPANE

The SPANE includes 12 items aimed to measure the frequency of positive (SPANE-P subscale) and negative (SPANE-N subscale) emotions, and it covers both general feelings (positive, good, pleasant, negative, bad, unpleasant) and specific emotions (happy, joyful, contented, sad, afraid, angry). As argued by the developers of the SPANE (Diener et al., 2010), the scale was developed to improve on existing measures of feelings and to overcome some of the well-known limitations and shortcomings of the PANAS. More specifically, the authors emphasize the following advantages of the SPANE: (1) the scale assesses a range of positive and negative feelings, because it covers both specific emotions and general feelings; (2) it covers a full range of arousal levels, in contrast to the PANAS, which covers predominantly high-arousal emotions; and (3) the items are framed in terms of the frequency of emotions and feelings (i.e., the amount of time the individual experiences each feeling and emotion), instead of intensity, which has weaker associations with other measures of well-being. In line with these advantages, the SPANE may be more appropriate for use across different cultural contexts, because it "can better reflect the full set of feelings felt by individuals around the globe" (Diener et al., 2010, p. 145). Although the expectation of cross-cultural consistency has been clearly voiced by the original authors—"the scale should perform well across societies" (Diener et al., 2010, p. 153)—no study to date has examined whether the SPANE operates similarly across various national samples.

The SPANE has gained increasing popularity in the field of well-being in the past decade and it has been used in hundreds of studies in at least 20 countries worldwide (Busseri, 2018). Moreover, it has been validated in more than a dozen countries, across different cultural contexts, such as China (F. Li et al., 2013), Germany (Rahm et al., 2017), Italy (Giuntoli et al., 2017), Japan (Sumi, 2014), Serbia (Jovanović et al., 2020), and South Africa (du Plessis & Guse, 2017). The correlated two-factor model of the SPANE has been largely supported in these studies, with strong correlations (typically above −.50) observed between the two factors (i.e., SPANE-P and SPANE-N). In addition, studies examining the psychometric properties of the SPANE provide evidence supporting the internal consistency of the SPANE-P and SPANE-N components (e.g., Daniel-González et al., 2020; Silva & Caetano, 2013), convergent validity (e.g., Kyriazos et al., 2018; Rahm et al., 2017), and measurement invariance across gender and age (e.g., Jovanović et al., 2020; Martin-Krumm et al., 2018).

However, despite promising psychometric properties found in both Western and non-Western societies, prior research has not yet investigated the cross-national consistency of the SPANE. As such, the cross-national invariance properties of the SPANE are still unknown. With the increasing popularity of cross-cultural studies on well-being, there is a clear need to evaluate which scales can be validly used to investigate levels, determinants, and outcomes of affective well-being across different countries.

The Present Study

The present study investigated the measurement invariance of the SPANE across 13 countries: China, Colombia, Germany, Greece, India, Italy, Japan, Poland, Portugal, Serbia, Spain, Turkey, and the United States, spanning four continents: Asia, Europe, North America, and South America. As shown in Table 1, these countries vary greatly across Hofstede's cultural dimensions (for details, see Hofstede et al., 2010), global indices of prosperity (Legatum Institute, 2020) and development (United Nations Development Programme, 2020), as well as on levels of happiness (Helliwell et al., 2020). For example, we included countries that can be considered highly collectivistic (e.g., China, Colombia, Serbia) or highly individualistic (e.g., Germany, the United States), countries that rank very high (e.g., Germany) or low (e.g., India) on measures of prosperity and development, and countries that differ considerably in average happiness (e.g., Germany ranked 17 and India ranked 144) over the past few years.

Furthermore, we included countries that differ in cultural construal of emotions and well-being. For example, individuals in East Asian and collectivist countries hold dialectical concepts of emotions, characterized by a balance between positive and negative emotions that are considered in the Western and individualistic societies to be opposites (Mitamura et al., 2014; Miyamoto & Ryff, 2011). These differences in cultural scripts are supported by findings showing that correlations between positive and negative emotional experiences are weaker among individuals from interdependent countries (e.g., India, Japan, and Russia) than among individuals from independent countries (e.g., Germany, UK, and the USA), with the former also

Table 1. Cultural Dimension Scores, Prosperity, and Well-Being Rankings for 13 Countries.

	Hofstede's cultural dimensions[a]						Prosperity and well-being ranking		
Country	Power Distance	Individualism	Masculinity	Uncertainty Avoidance	Long-Term Orientation	Indulgence	LPI rank[b] 2020	HDI rank[c] 2019	WHR rank[d] 2017-2019
China	80	20	66	30	87	24	54	85	94
Colombia	67	13	64	80	13	83	74	83	44
Germany	35	67	66	65	83	40	8	6	17
Greece	60	35	57	100	45	50	41	32	77
India	77	48	56	40	51	26	101	131	144
Italy	50	76	70	75	61	30	31	29	30
Japan	54	46	95	92	88	42	19	19	62
Poland	68	60	64	93	38	29	36	35	43
Portugal	63	27	31	99	28	33	27	38	59
Serbia	86	25	43	92	52	28	52	64	64
Spain	57	51	42	86	48	44	24	25	28
Turkey	66	37	45	85	46	49	94	54	93
United States	40	91	62	46	26	68	18	17	18

Note. LPI = Legatum Prosperity Index; HDI = Human Development Index; WHR = World Happiness Report. [a]Values were retrieved from https://www.hofstede-insights.com/product/compare-countries. Scores on each dimension range from 0 to 100. [b]Ranks range from 1 to 167. [c]Ranks range from 1 to 189. [d]Ranks range from 1 to 156.

exhibiting a greater tendency to experience positive and negative emotions simultaneously (Grossmann et al., 2016). In addition, countries included in the present study differ in conceptualizations of happiness as portrayed in dictionary definitions of happiness (Oishi et al., 2013). For example, whereas German, Italian, Japanese, Portuguese, and Turkish definitions of happiness include luck and fortune (i.e., favorable external circumstances), these features are absent from definitions in the United States, Spain, and India, which center on positive inner states (Oishi et al., 2013). Differences in the conceptualizations, cultural scripts, and norms of positive and negative emotional experience and well-being across countries call for research that would carefully examine the cross-cultural consistency of the SPANE and evaluate whether cross-country comparisons based on this scale are justified. Thus, the main goal of the present study was to investigate whether the scale operates similarly across different languages and cultural contexts, and enables meaningful cross-country comparison of positive and negative emotional experiences, by evaluating the measurement invariance of the SPANE across 13 countries. We believe that the inclusion of both Western and non-Western countries, which vary in cultural values and construal of emotions, enables a robust test of the cross-national measurement invariance of the SPANE.

Method

Sample and Procedure

The present study used adult samples (age range 18-85 years; 58.2% female) from 13 countries. Sample sizes ranged from 546 (India) to 2,264 (Greece), and the total sample size is 12,635. Demographic characteristics for each country are shown in Table 2. A detailed description of the sample and procedure for each country is given below.

Chinese Sample. The Chinese sample consisted of 995 participants (54.5% female), with a mean age of 40.04 years (SD = 16.67; age range = 18-85 years). The majority of the participants were married (59.0%; 39.3% single; 0.3% divorced; 0.9% widowed; 0.5% did not answer), were employed (65.4%; 4.7% unemployed; 10.6% retired; 12.5% student; 6.7% homemaker; 0.1% did not answer), and had secondary level education (45.2%; 41.0% tertiary level education; 10.9% primary level education; 2.7% did not complete primary level education; 0.2% did not answer). The data were collected in 2014 by phone. The telephone number was randomly drawn from the local telephone directory, and eligible participants were selected based on the last birthday rule. Verbal consent was obtained from all participants, and participants did not receive any compensation for their participation. The data from the Chinese sample has been used in another study (Tong & Wang, 2017).

Colombian Sample. The Colombian sample was composed of 1,240 participants (64.4% female), with a mean age of 25.63 years (SD = 8.63; age range 18-67 years). With regard to education, 42.6% had completed at least college/university education, 41.4% had completed high school level, and 15.9% had finished compulsory secondary studies or had only primary studies. Regarding marital status, 75.5% were single, 22.0% were married or cohabiting, and 2.5% were divorced or widowed. Regarding working status, 43.8% were

Table 2. Demographic Characteristics of the Sample by Country.

Country	Language used	Administration mode	Sample size	% Female	Age range	Mean age (*SD*)
China	Chinese	Telephone	995	54.5	18-85	40.04 (16.67)
Colombia	Spanish	Online	1,240	64.4	18-67	25.63 (8.63)
Germany	German	Online	902	77.4	18-80	25.60 (7.11)
Greece	Greek	Online	2,264	63.3	18-69	35.61 (12.32)
India	Hindi	Paper-and-pencil	546	39.4	18-59	28.43 (7.51)
Italy	Italian	Online	706	72.8	18-76	30.85 (10.36)
Japan	Japanese	Paper-and-pencil	749	49.0	20-68	42.20 (12.48)
Poland	Polish	Online	1,054	52.4	18-83	44.16 (15.26)
Portugal	Portuguese	Online	741	48.7	25-52	37.23 (7.43)
Serbia	Serbian	Online	1,214	54.9	20-82	43.41 (11.42)
Spain	Spanish	Online	981	69.0	18-71	28.40 (11.47)
Turkey	Turkish	Online	642	53.6	18-63	31.15 (13.43)
United States	English	Online	601	38.6	19-72	35.44 (9.84)

students only, 26.2% were students and had sporadic or part-time jobs, 23.6% were employed or self-employed, 5.0% were unemployed, 1.0% were inactive, and 0.4% were retired. Participants were recruited during 2019-2020 (the data collection was completed at the beginning of March 2020, before the outbreak of the COVID-19 pandemic in Colombia) by different means (email, social networks, and direct solicitation). Data were collected online with LimeSurvey, an open-source survey tool. Participants were included if they identified themselves as Colombians in the survey and if they were at least 18 years of age. Participants were presented with an explanation of the study and then provided informed consent prior to completing the survey. The study was approved by the Ethics Committee of the Cooperative University of Colombia, which guarantees that data collection complied with the Colombian Law of Data, ensuring confidentiality and anonymity. The participants did not receive any compensation for participating in the study. The data from the Colombian sample were used in previously published studies (Martín-Carbonell et al., 2021; Martín-Carbonell et al., in press).

German Sample. The German sample consisted of 902 participants (77.4% female), with a mean age of 25.60 years ($SD = 7.11$; age range = 18-80 years). The majority of the participants were not married (89.3%; 8.1% married; 1.5% divorced; 1.1% separated; 0% widowed), were in a romantic relationship (61.3%; 38.7% not in a romantic relationship), were students (69.4%; 19.4% employed, 3.0% self-employed; 1.9% unemployed; 6.3% other), and indicated having a general higher education entrance qualification (German Abitur; 86.0%; 4.6% secondary level education; 5.3% technical secondary school; 1.3% still in school; 2.8% other). The German data originated from three different studies. One data set (54.0%) was derived from the baseline assessment of an experience sampling study that was collected in 2018. Participants could receive up to €20 as compensation for participation. Parts of the data have already been analyzed (Krasko et al., 2020). Another data set (14.1%) was derived from the baseline assessment of another experience sampling study that was collected in 2019 (Krasko et al., 2019). To date, these data have only been analyzed for two master theses. Data from both experience sampling studies were collected using the mobile phone application ExpiWell. The third data set (31.9%) was derived from an online survey using Qualtrics and was collected in 2017. Parts of the data have already been analyzed (Krasko et al., 2021). For all aforementioned studies, participants were recruited using different online (e.g., Facebook groups) and offline (e.g., flyers in supermarkets) sources. Informed consent was obtained from all participants included in the studies. For compensation, participants could participate in lotteries of different vouchers in all studies mentioned above. Student participants could receive course credits (instead of financial compensation when this applied).

Greek Sample. The Greek sample involved 2,264 participants (63.3% female) with a mean age of 35.61 years ($SD = 12.32$, age range = 18-69 years). The majority of the participants were single (51.1%; 41.1% married/living together; 7.8% divorced). Most participants were employed (40.5%; 36.9% housekeepers; 16.4% retired; 2.4% unemployed; 3.8% other) with a tertiary level education (73.8%; 25.6% secondary level education; 0.6% primary level education). The questionnaire was administered online, and informed consent was obtained. Data were collected with the network sampling method. Specifically, each psychology student voluntarily recruited at least 15 adult (nonstudent) participants from their social environment during 2017-2018, and received extra course credit. Recruitment rules allowed students to recruit participants from their

social environment, without taking the questionnaire themselves. The data from the Greek sample were used in a previous study (Kyriazos et al., 2018).

Indian Sample. The Indian sample consisted of 546 participants (39.4% female), with a mean age of 28.43 years (SD = 7.51; age range = 18-59 years). The majority of the participants were married (48.5%; 43.2% single; 0.5% divorced; 0.2% widowed; 7.5% did not answer), employed (64.1%; 23.5% student, 0.4% unemployed; 4.5% homemakers; 7.5% did not answer), and had tertiary level education (84.0%; 13.4% secondary level education; 2.2% primary level education; 0.4% did not complete any education). A data booklet comprising the paper-and-pencil version of a number of psychological measures, including the SPANE, along with a demographic information sheet and informed consent form, were distributed to a convenience sample of Indians between 2013 and 2014. Informed consent was obtained from all the participants, and they did not receive any compensation for their research participation. The data were used in a previous study in which the SPANE was used as a validation measure for assessing the concurrent validity of some newly developed well-being measures in the Indian context (Singh et al., 2016).

Italian Sample. The Italian sample consisted of 706 participants (72.8% female), with a mean age of 30.85 years (SD = 10.36; age range = 18-76 years). The sample included 39.9% unemployed participants; 30.6% were employed, and 23.8% were students (5.7% reported "other"). Participants were recruited online via social networking websites and through email requests in 2016. All participants were informed through written instructions about the anonymity and confidentiality of their data; participants did not receive any compensation for participating in the study. The data from the Italian sample were used in a previous study (Giuntoli et al., 2017).

Japanese Sample. The Japanese sample consisted of 749 participants (49.0% female), with a mean age of 42.20 years (SD = 12.48; age range = 20-68 years). They were employed in a variety of occupations, which mainly included clerical (39.8%), factory (30.3%), and sales work (9.3%). Most of the participants (92.7%) worked in urban areas in Japan. The data were collected from a nonrandom sample of Japanese adults attending several workshops or lecture meetings (who completed the questionnaire at the workshop or lecture meeting) and their coworkers who completed the questionnaire in their offices from 2014 to 2018. The questionnaire was administered in paper-and-pencil format. The participants voluntarily participated in the study without compensation and provided informed consent. The data from this sample were not used in previously published studies.

Polish Sample. The Polish (representative) sample consisted of 1,054 participants (52.4% female), with a mean age of 44.16 years (SD = 15.26; age range = 18-83 years). The highest number of the participants had secondary level education (48.2%; 40.5% tertiary level education; 1.8% primary level education; 9.5% vocational education). The questionnaire was administered online and the data were collected by the ARIADNA research company from preregistered users of an online research panel. Informed consent was obtained from all participants included in the study; participants received compensation for participating in the study in the form of loyalty points (that can be later exchanged for rewards, like cups, books, or tickets). The data regarding SPANE from the Polish sample were not used in previous studies.

Portuguese Sample. The Portuguese sample included 741 participants (48.7% female), with a mean age of 37.23 years (SD = 7.43; age range = 25-52 years). The majority of the participants were married (49.5%; 36.2% single; 5.8% in a romantic relationship; 8.5% divorced or widowed). A large proportion of the participants were full-time employed (82.2%), 6.3% were part-time employed, and 11.5% reported an "other" work-related situation. Similar proportions of participants reported completed secondary education (42.0%) or holding a university degree (41.9%), followed by smaller rates of individuals with less than the secondary level education (10.8%) or a master's degree (5.3%). The questionnaire was administered online, and the data were collected using a snowball sampling technique between 2011 and 2012. Informed consent was obtained from all participants included, and they did not receive any compensation for participating in the study. The data from the Portuguese sample were used in a previous study (Silva & Caetano, 2013).

Serbian Sample. The Serbian sample consisted of 1,214 participants (54.9% female), with a mean age of 43.41 years (SD = 11.42; age range = 20-82 years). The majority of the participants were married (63.8%; 15.7% single; 11.4% in a romantic relationship; 5.1% divorced; 3.7% widowed; 0.3% did not answer), were employed (77.2%; 12.4% unemployed; 9.1% retired; 1.3% did not answer), and had tertiary level education (53.1%; 44.8% secondary level education; 1.8% primary level education; 0.3% did not answer). The questionnaire was administered online, and the data were collected using a snowball sampling technique during 2018 and 2019. Informed consent was obtained from all participants included in the study, and the participants did not receive any compensation for participating in the study. The data from the Serbian sample were not used in previous studies.

Spanish Sample. The Spanish sample was composed of 981 participants (69.0% female), with a mean age of

28.40 years (SD = 11.47; age range = 18-71 years). The majority of the participants completed high school (55.2%) or college (41.0%; 3.8% did not complete high school) and were single (63.0%; 34.0% married or cohabiting; 3.0% divorced or widowed). Regarding working status, 41.7% were students only, 25.0% were students and had sporadic or part-time jobs, 27.4% were employed or self-employed, 3.7% were unemployed, 0.9% were inactive, and 1.2% were retired. Participants were recruited in 2018 via email and social networks including an explanation of the study and a link to LimeSurvey. They had to read and accept an online informed consent before entering the study. Participants were included if they identified themselves as Spaniards in the survey and if they were at least 18 years of age. The study was conducted in compliance with Spanish legislation (Ley Orgánica 3/2018, December 5) and the code of ethics for research involving human subjects, as outlined by the Universitat de València Human Research Ethics Committee. The participants did not receive any compensation for participating in the study. The data from the Spanish sample have been used in a previous study on the psychometric properties of the SPANE (Espejo et al., 2020).

Turkish Sample. The Turkish sample included 642 participants (53.6% female), with a mean age of 31.15 years (SD = 13.43; age range = 18-63 years). The questionnaire was administered online, and the data were collected using convenience sampling during 2018 and 2020 (during February 2020, before the outbreak of the COVID-19 pandemic in Turkey). Informed consent was obtained from all participants included in the study, and the participants did not receive any compensation for participating. A part of the data from the Turkish sample (n = 256) was reported in Yildirim (2019), while the remaining part was not used in previous studies.

U.S. Sample. The U.S. sample consisted of 601 participants (38.6% female), with a mean age of 35.44 (SD = 9.84; age range = 19-72 years). The majority of the participants were never married (48.9%; 42.4% married; 7.5% divorced; 0.8% separated; 0.3% widowed) and had at least some tertiary level education (87.1%; 12.9% secondary education). All participants were employed (88.4% full-time). Participants were recruited via Amazon Mechanical Turk (MTurk), and the questionnaire was completed online using Qualtrics in June 2019. Only high reputation workers (i.e., those with at least 95% of their past tasks approved on MTurk) were allowed to participate in order to maximize data quality (Peer et al., 2014). Seven hundred twenty-two individuals attempted to participate, but 121 participants (16.8%) did not meet inclusion criteria (living in the United States, English-speaking, and currently employed outside of MTurk) or failed at least 3/5 attention checks. Informed consent was obtained from all participants included in the study. Participants were compensated US$3 for completing the survey. The data from the U.S. sample were used as part of the author's doctoral dissertation (Rice, 2020a).

Measures

SPANE (Diener et al., 2010) consists of 12 items. Items are rated on a 5-point scale ranging from 1 (*very rarely or never*) to 5 (*very often or always*). The time frame used in the SPANE is "the past four weeks." The official SPANE versions for each country were used (see online Table S1 for the SPANE translation in each country). The back-translation approach was used for each country's surveys following general guidelines for cross-cultural adaptation and translation of psychological instruments (Epstein et al., 2015). Descriptive statistics and internal consistency reliability of the SPANE-P and the SPANE-N in each country are presented in Table S2 online. Both SPANE-P and SPANE-N subscales demonstrated adequate internal consistency in each country: Cronbach's α for SPANE-P ranged from .80 in India to .93 in Italy, whereas α for SPANE-N ranged from .78 in Germany to .93 in the United States.

Statistical Analysis

Data were analyzed using M*plus* Version 8.4 (Muthén & Muthén, 1998-2017). Single-factor and two-factor confirmatory factor analysis (CFA) models were tested separately in each nation. Given that the variables are ordered categorical, the WLSMV (mean and variance-adjusted weighted least squares) estimator with a PROBIT link function was used. WLSMV is widely used with categorical data (C. Li, 2016), and it has been shown to perform better than maximum likelihood when five response options are used (Beauducel & Herzberg, 2006). Cases with missing values on all items were removed from the analysis (N = 29). There were also 104 participants (0.8% of the total sample) with missing values on one or more items. With categorical outcomes and no covariates in the model, robust weighted least squares estimators in M*plus* utilize all available data to calculate each correlation (Asparouhov & Muthen, 2010). Notably, with such a small number of missing values, various missing data management strategies are not expected to influence the results in significant ways (Tabachnick & Fidell, 2015).

A minimum cutoff of .95 for the comparative fit index (CFI), a maximum cutoff of .07 for the root mean square error of approximation (RMSEA), and a maximum cutoff of .08 for the standard root mean square residual (SRMR) were considered as indicative of good fit (e.g., Kline, 2016). Values between .07 and .10 for RMSEA suggest a mediocre fit. As shown in the Results section, the two-factor model in this study had CFI and SRMR values suggesting a good fit

in all countries, yet RMSEA values were indicative of a bad fit (values over .10) in three countries. Accordingly, we modified the models only if RMSEA values were over .10.

Model-based reliability and convergent validity analyses of the SPANE latent variables were performed next. First, the factor measurement reliability (Kline, 2016) of the specified SPANE solution for each country was evaluated with composite reliability (CR; Werts et al., 1974), which is identical to the ω coefficient (McDonald, 1999), because we used the standardized factor loadings (Raykov, 1997; but see also Kline, 2016). Then, the average variance extracted (AVE; Fornell & Larcker, 1981) was estimated for each country to examine factor measurement validity. In addition, the internal consistency reliability of the SPANE was calculated in each country, using Cronbach's alpha (Cronbach, 1951), its 95% confidence interval (Feldt et al., 1987), and its greatest lower bound estimate (glb; Jackson & Agunwamba, 1977). Generally, a value of Cronbach's alpha $\geq .70$ is considered acceptable, and the same is true for CR values (Hair et al., 2010). Values of AVE $\geq .50$ are considered adequate (Fornell & Larcker, 1981).

Measurement invariance was tested under the framework of multiple-group CFA with delta parameterization (Millsap, 2012; Millsap & Yun-Tein, 2004). Three types of measurement invariance were of interest: configural invariance (identical factor structures; i.e., the same number of factors and items and the same patterns of free and fixed loadings), metric invariance (equality of factor loadings), and scalar invariance (equality of factor loadings and thresholds). A configural model was first tested as a baseline model. In this model, all factor loadings and thresholds were estimated freely across countries. Unlike in models with continuous indicators, in models with categorical indicators with delta parameterization, metric invariance cannot be tested separately from scalar invariance (Muthén & Muthén, 1998-2017; Wu & Estabrook, 2016). Thus, a scalar invariance model was tested where equality constraints were simultaneously imposed on factor loadings and thresholds. Measurement invariance was examined by comparing the fit indices of the configural model and those of the scalar model. We used the cutoff criteria conventionally used to indicate scalar invariance: $\Delta CFI \leq -.010$ and $\Delta RMSEA \leq .015$ (Chen, 2007). When full measurement invariance is not supported, partial measurement invariance may be an alternative to explore (Byrne et al., 1989; Wang & Wang, 2019). In such models, whereas most of the model parameters remain constrained to be invariant across groups, a subset of parameters (e.g., some factor loadings and thresholds) are allowed to vary across groups. If partial measurement invariance holds, meaningful comparisons of latent variables across groups can be performed. To establish partial measurement invariance, loadings and thresholds should be invariant for at least two items per latent factor (Byrne et al., 1989), or if using more conservative criteria, at least half of the indicators (in the case of the SPANE, three items per factor) (Vandenberg & Lance, 2000).

Results

Sociodemographic Comparisons

As indicated in Table 2, there were a wide range of sociodemographic characteristics across the different country samples. Indeed, age, Kruskal–Wallis $H(12) = 3287.03$, $p < .01$, and gender, $\chi^2(4, N = 12{,}616) = 581.62$, $p < .01$, were significantly different between countries. Additionally, characteristics such as marital status, education level, and employment type also notably varied. However, direct comparisons were not feasible due to distinct types of data collected (e.g., no data on marital status were collected from Turkish participants) and inconsistent survey response options (e.g., Spanish participants had the option to select "student," "retired," "unemployed," etc., as employment status, whereas being employed was an inclusion criterion for participating in the U.S. data collection, so only job type was recorded). Additionally, the sample in 10 countries ($n = 10{,}345$; 81.9%) provided data via online surveys, those in two countries used paper-and-pencil ($n = 1{,}295$; 10.2%), and those in one country used a telephone survey ($n = 995$; 7.9%).

Single-Group CFAs

As shown in Table 3, whereas the single-factor models did not fit the data well, the two-factor models provided a considerably better fit. This result indicates that negative and positive emotion items form two distinct factors, rather than a single factor, in all countries. Therefore, single-factor models were excluded from the subsequent analyses. The correlated two-factor model provided an acceptable fit in 10 of the countries. The RMSEA values were greater than the cutoff value of .10 in Colombia (.103, 90% CI [.096, .109]), Spain (.111, 90% CI [.104, .118]), and Turkey (.126, 90% CI [.117, .135]). Thus, the models were modified in these countries by adding item residual covariances based on modification indices. Results of these modifications are shown in Table 3. Adding one residual covariance in Colombia and Spain (between items "good" and "bad") and three in Turkey (positive-negative; good-positive; bad-negative) resulted in RMSEA values $< .10$ in these countries. These residual covariances are theoretically justifiable, because they refer to pairs of items capturing general feelings that are either opposite ("good" and "bad," "positive" and "negative") or same in valence ("good" and "positive," "bad" and "negative"). General feelings are saturated with the valence dimension (pleasure/displeasure) of the emotion circumplex (Diener et al., 2010) and have similar arousal levels. Furthermore, they refer to concepts that are semantically

Table 3. Results of the Single-Group CFA Models in 13 Countries.

	Single-factor model				Two-factor model with correlated factors					
Country	χ^{2a}	RMSEA [90% CI]	CFI	SRMR	χ^2	*df*	RMSEA [90% CI]	CFI	SRMR	Items residual covariance
China	2606.442	.218 [.211, .225]	.704	.125	385.794	53	.079 [.072, .087]	.961	.041	—
Colombia	1788.719	.161 [.155, .167]	.872	.080	616.025	52	.094 [.087, .100]	.959	.038	3 and 4
Germany	628.540	.109 [.101, .116]	.951	.045	250.475	53	.064 [.056, .072]	.983	.030	—
Greece	3890.087	.177 [.172, .182]	.930	.068	947.531	53	.086 [.082, .091]	.984	.026	—
India	2077.249	.262 [.252, .272]	.534	.167	241.151	53	.081 [.070, .091]	.957	.047	—
Italy	1406.914	.188 [.180, .197]	.931	.083	405.281	53	.097 [.088, .106]	.982	.029	—
Japan	2426.313	.242 [.234, .250]	.850	.170	193.843	53	.060 [.051, .069]	.991	.028	—
Poland	3340.218	.240 [.233, .247]	.877	.128	573.593	53	.097 [.089, .104]	.981	.032	—
Portugal	1544.464	.193 [.185, .201]	.885	.111	337.055	53	.085 [.076, .094]	.978	.037	—
Serbia	1943.488	.170 [.163, .176]	.930	.068	506.997	53	.084 [.077, .091]	.983	.029	—
Spain	1157.918	.144 [.137, .152]	.891	.069	542.700	52	.098 [.091, .106]	.952	.039	3 and 4
Turkey	1149.724	.178 [.169, .187]	.924	.061	359.204	50	.098 [.089, .108]	.979	.032	1 and 3, 2 and 4, 1 and 2
USA	2504.693	.275 [.266, .284]	.910	.119	312.572	53	.090 [.081, .100]	.991	.035	—

Note. *df* = degrees of freedom; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; SRMR = standardized root mean square residual. Item numbers correspond to the following terms: 1: "positive"; 2: "negative"; 3: "good"; 4: "bad."
[a]There are 54 *df* for entire column.

close (e.g., good/positive, bad/negative) or antonyms (e.g., good/bad, positive/negative). Therefore, at least in some cultures, they are expected to be more closely interrelated than SPANE's specific emotions (viz., happy, joyful, contented, sad, afraid, and angry) that differ in their arousal levels and capture more conceptually distinct emotional experiences (Scherer et al., 2013; Yik et al., 2011). These residual covariances were included for Colombia, Spain, and Turkey, in the following invariance models.

For the factor measurement reliability (Kline, 2016) of the two-factor SPANE solution across countries, the CR for SPANE-P ranged from .84 (India) to .95 (Italy). The CR for SPANE-N ranged from .82 (Germany and Spain) to .95 (the United States). The AVE ranged from .48 (India) to .76 (Italy) for SPANE-P and from .44 (Germany) to .75 (the United States) for SPANE-N. The CR and AVE estimates are reported in Table 4.

The internal consistency reliability for SPANE-P ranged from α = .80, 95% CI [.77, .83] (India) to .93, 95% CI [.92, .94] (Italy). For SPANE-N, it ranged from α = .78, 95% CI [.76, .80] (Germany) to .93, 95% CI [.92, .94] (the United States). The glb estimate for SPANE-P ranged from .83 (Spain) to .94 (Poland, Japan, and USA). The glb for SPANE-N ranged from .79 (Germany and Spain) to .93 (the United States). SPANE-P's glb was less than the corresponding alpha for Italy, Portugal, and Spain. The alpha coefficients and glb estimates are presented in the online Table S2.

Table 4. Factor Measurement Reliability (CR) and Factor Measurement Convergent Validity (AVE).

	SPANE-P		SPANE-N	
Country	CR	AVE	CR	AVE
China	.86	.50	.86	.51
Colombia	.89	.58	.84	.47
Germany	.89	.57	.82	.44
Greece	.93	.68	.89	.58
India	.84	.48	.86	.51
Italy	.95	.76	.89	.58
Japan	.94	.73	.90	.61
Poland	.94	.74	.91	.64
Portugal	.93	.69	.87	.54
Serbia	.94	.71	.87	.54
Spain	.89	.59	.82	.45
Turkey	.92	.65	.84	.48
USA	.94	.73	.95	.75

Note. CR = composite reliability or factor rho coefficient (Kline, 2016); AVE = average variance extracted.

Multigroup Confirmatory Factor Analysis

With all the models demonstrating acceptable fit and reliability, a multiple-group configural invariance model was specified with no constraints on factor loadings and thresholds (see Table 5 for standardized factor loadings). As presented in Table 6, the fit indices of this model were acceptable, indicating that the two latent variables can invariably be measured across the countries using a two-factor structure as specified in this analysis. To evaluate metric and scalar invariance, equality constraints were imposed on the factor loadings and thresholds simultaneously in a separate model. As shown in Table 6, comparing the fit indices of this model with those of the configural model resulted in a ΔRMSEA = .013 and ΔCFI = −.024. Because the ΔCFI was greater than our a priori threshold, we proceeded with

Table 5. Standardized Factor Loadings from the Configural Model of the SPANE.

	Standardized factor loading												
Item	China	Colombia	Germany	Greece	India	Italy	Japan	Poland	Portugal	Serbia	Spain	Turkey	USA
SPANE-P													
1. positive	.57	.79	.83	.89	.67	.91	.78	.88	.78	.85	.84	.80	.90
3. good	.69	.67	.83	.86	.81	.79	.84	.88	.83	.82	.66	.75	.92
5. pleasant	.65	.74	.74	.87	.69	.87	.88	.83	.74	.83	.72	.88	.90
7. happy	.80	.84	.77	.82	.73	.93	.91	.87	.90	.88	.84	.91	.92
10. joyful	.78	.83	.48	.71	.67	.88	.87	.82	.85	.86	.79	.75	.76
12. contented	.73	.68	.82	.78	.55	.84	.84	.87	.87	.82	.72	.74	.70
SPANE-N													
2. negative	.73	.77	.78	.86	.74	.93	.85	.87	.85	.82	.86	.74	.93
4. bad	.81	.72	.80	.91	.77	.72	.83	.90	.80	.86	.62	.73	.94
6. unpleasant	.76	.72	.67	.88	.81	.80	.86	.85	.69	.75	.72	.88	.93
8. sad	.78	.76	.74	.80	.76	.82	.72	.83	.82	.76	.79	.81	.89
9. afraid	.58	.49	.48	.47	.66	.58	.72	.65	.57	.57	.44	.35	.72
11. angry	.57	.59	.43	.51	.49	.68	.67	.64	.65	.60	.47	.54	.77

Note. SPANE = Scale of Positive and Negative Experience.

Table 6. Results of the Multigroup CFA Across 13 Countries.

Model	χ^2	*df*	RMSEA [90% CI]	CFI	SRMR	ΔRMSEA	ΔCFI
Configural	5284.596	684	.083 [.081, .085]	.982	.034	—	—
Full Scalar	12119.892	1212	.096 [.095, .098]	.958	.044	.013	−.024
Partial Scalar (Item 9[a])	9726.444	1164	.087 [.085, .089]	.967	.041	.004	−.015
Partial Scalar (Items 9 and 11[b])	9002.333	1116	.085 [.084, .087]	.970	.040	.002	−.012
Partial Scalar (Items 9, 11, and 8[c])	8211.075	1068	.083 [.081, .085]	.973	.039	.000	−.009

Note. df = degrees of freedom; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; SRMR = standardized root mean square residual; ΔRMSEA = difference in RMSEA between compared models; ΔCFI = difference in CFI between compared models. All chi-square coefficients are significant at $p < .05$.
[a]Item 9 is "afraid". [b]Item 11 is "angry." [c]Item 8 is "sad."

testing a sequence of partial measurement invariance models, in every one of which equality constraints on the factor loadings and thresholds of one additional item were relaxed. The order in which the constraints were released depended on the sizes of the modification indices across the groups. We started with a model where the equality constraints on the factor loadings and thresholds of Item 9 ("afraid") were relaxed across all countries. As shown in Table 6, this model had a better fit than the full invariance model, yet the ΔCFI was still greater than −0.010. Another model with relaxed constraints on Items 9 and 11 ("angry") also did not satisfy the fit requirement. Additionally freeing Item 8 ("sad") in a final model, however, resulted in a ΔCFI that was smaller than −.010. Thus, our results suggested that full measurement invariance held for positive affect, and partial measurement invariance held for negative affect. Items 9, 11, and 8 (all measuring specific negative emotions) were not invariant across the countries included in this analysis.

Latent correlations between SPANE-P and SPANE-N ranged from −.11 in India to −.82 in Germany (see online Table S3). Estimation of latent means (online Table S3) showed that the highest levels of positive emotions (SPANE-P) were reported in Colombia and Portugal, and lowest levels in Italy and Turkey, whereas the highest levels of negative emotions (SPANE-N) were reported in Turkey and India, and the lowest were reported in China and Colombia.

Discussion

The present study investigated the cross-national measurement invariance of the SPANE across 13 countries to evaluate whether the use of this scale for cross-cultural comparisons of affective experience is justified in different cultural contexts. The correlated two-factor model of the SPANE provided an acceptable fit in most countries, except Colombia, Spain, and Turkey, in which some correlated residuals between pairs of opposite items (viz., "good" and "bad" in Spain and Colombia; "positive" and "negative" in Turkey) or same valence items ("good" and "positive," and

"bad" and "negative" in Turkey) had to be added to achieve an adequate fit.

Latent correlations between SPANE-P and SPANE-N varied remarkably across countries, with correlations ranging from −.11 in India to −.82 in Germany. The weakest negative correlations between the two dimensions were found in three Asian countries (China, India, and Japan), and these correlations were substantially weaker than those found in the remaining 10 countries. These results reveal cultural differences in the relationship between positive and negative emotions, which have been consistently found in previous studies using different measures of emotions. For example, the co-occurrence of positive and negative emotions is more prevalent in East Asian than in Western cultures (e.g., Schimmack et al., 2002), and whereas positive and negative emotions are typically significantly negatively correlated in Western samples (e.g., Crawford & Henry, 2004; Thompson, 2007), the correlation between positive and negative emotions in East Asian samples is often found to be weak or nonsignificant (Liu et al., 2020; Spencer-Rodgers et al., 2010) or even positive (Bagozzi et al., 1999). Furthermore, Chen et al. (2016) found that positive and negative affect were more strongly related in the United States than in China across all levels of activation (low, moderate, and high), supporting the greater independence of the two affective dimensions in East Asia. In sum, the pattern of correlations between positive and negative emotional experiences across countries observed in the present study supports the notion that a dialectical understanding of emotions is more strongly endorsed in Eastern than in Western cultural contexts (Wilken & Miyamoto, 2018).

Factor measurement reliability and validity, as well as the internal consistency reliability were satisfactory and comparable across countries. Configural invariance was supported for the SPANE, suggesting that the two dimensions of emotional experiences are similar across countries. However, full scalar invariance was not supported, as three items capturing specific negative emotions (sad, afraid, and angry) were found to operate differently across countries. After relaxing loadings and thresholds of these three items, there was evidence for partial measurement invariance. Cross-cultural variation in the meaning and subjective experience of emotions is well documented (e.g., Scollon et al., 2011), and terms used to describe a particular emotion may have a different meaning across cultures. In the analysis of 2,474 languages across 20 language families, Jackson et al. (2019) found high variability in the meanings of emotion terms across cultures, and different patterns of associations between emotion concepts across different language families. For example, "fear" was closely associated with "surprised" among Austronesian and Austroasiatic languages, but was more closely associated with "anxiety" among Tai-Kadai, Nakh-Daghestanian, and Indo-European languages. In addition, "anger" was closely related to "envy" among Nakh-Daghestanian languages, but was more closely associated with "hate," "bad," and "proud" among Austronesian languages. Furthermore, Thompson et al. (2020) investigated semantic alignment of 1,010 concepts (across 21 semantic domains, including emotions) in 41 languages and found that the meaning of words related to emotions, such as sad, varied greatly across languages. Although the universality of concepts descriptive of basic emotions (such as afraid, angry, happy, satisfied) has been supported in the analysis of mutually isolated languages (Saucier et al., 2014), the findings of Jackson et al. (2019) and Thompson et al. (2020) indicate that the meaning of emotion terms might differ across languages, and caution is needed when aiming to compare self-reported emotions in cross-cultural studies. Our findings suggest that this caution should be exercised especially when aiming to compare specific negative emotions as measured with the SPANE across cultures. It is important to note that Kööts-Ausmees et al. (2013) found that negative emotion items (depressed, lonely, sad, and anxious) were invariant between 21 European countries. This suggests that (non)invariance of the items used to measure negative emotions might be sample specific, and that testing for measurement invariance is a necessary prerequisite before meaningful cross-national comparisons are made. Noninvariance of the three specific negative emotion items found in the present study might have been determined in part by emotions used in the SPANE and the countries included, so future research that will examine invariance of a range of emotions across a diverse set of languages and cultural settings is needed.

In contrast, our results suggest that terms descriptive of both general and specific positive emotions have similar meanings across languages and samples used in the present study. As research on measurement invariance of the scales measuring emotions is scarce, and studies on positive emotions still lag behind those of negative emotions, no definite conclusions on the invariance of positive emotion terms should be made. Furthermore, the majority of previous studies that examined measurement invariance of positive and negative emotions focused on only two languages (e.g., Lee et al., 2020) and produced inconsistent findings. For example, Davis et al. (2020) examined measurement invariance of the PANAS across American and Arab students, and found that metric invariance was established for all items except four negative affect items (irritable, nervous, scared, and jittery), whereas scalar invariance was supported for seven positive affect items (active, alert, attentive, determined, excited, inspired, strong) and only two negative affect items (distressed, hostile). On the other hand, another study that investigated measurement invariance of the Mroczek and Kolarz's (1998) positive and negative affect scales across Iran and the United States (Joshanloo & Bakhshi, 2016) found evidence for invariance of all items except two positive emotion items ("in good spirits" and

"full of life"). In sum, these findings clearly indicate that both differences and similarities in the meaning of emotion terms might be expected in the research on self-reported emotions, depending on the measures used and countries included in the study.

This was the first study to examine the measurement invariance of the SPANE across a large number of countries. The study has some notable strengths. First, we provided a strong contribution to increasing the scope and application of the SPANE worldwide, which is particularly important given recent arguments that this scale may be more theoretically aligned with the measurement of SWB and more strongly related to common well-being correlates than other common affect scales (Busseri, 2018; Rice & Shorey-Fennell, 2020). In addition, we utilized relatively large sample sizes from culturally diverse nations across multiple continents, increasing the generalizability of our findings. Finally, we conducted a robust statistical analysis, taking into account the categorical nature of the items, and best practices for assessing invariance with ordinal data. However, several limitations need to be noted.

First, although we included adult samples from both Western and non-Western countries spanning four continents, future research should seek to examine the measurement invariance of the SPANE in even more diverse cultures and across languages around the world, including countries from understudied regions such as Africa and the Middle East. Second, the samples included in our study were not nationally representative (except for Poland) and they varied in total size and sociodemographic characteristics (e.g., mean age, gender ratio). Simulation studies have shown that severely unbalanced sample size across groups might affect measurement invariance testing results (e.g., Yoon & Lai, 2018). Significant differences in sociodemographic variables between country samples further reduce the generalizability of our findings. However, previous research has established measurement invariance of the SPANE across characteristics such as age and/or gender in China (F. Li et al., 2013), Greece (Kyriazos et al., 2018), Italy (Giuntoli et al., 2017), Serbia (Jovanović et al., 2020), Spain (Espejo et al., 2020), and the United States (Rice, 2020b). Additionally, at least one analysis supported the invariance of the SPANE between online and pen-and-paper administrations (Giuntoli et al., 2017). As such, the potential confounding effects of these demographic discrepancies are likely negligible in the present study. Nonetheless, future cross-national studies of the SPANE should aim to use more balanced country samples, and future research may benefit from further measurement invariance evaluations across additional, less studied sociodemographic characteristics (e.g., relationship status, education level), which are yet to be fully understood. Third, cutoff recommendations for invariance vary from study to study, based on a wide variety of factors (e.g., variable type and sample size). Although we relied on Chen's (2007) recommendations, more stringent criteria (e.g., ΔCFI of .002; Meade et al., 2008) and alternative procedures (e.g., avoiding CFI to evaluate invariance; Kang et al., 2016) have been suggested, the use of which may yield different results from those found in the present study. Fourth, the mode of administration (online, paper-and-pencil, telephone interviewing) was not uniform in all countries. Although previous studies found that the effect of administration modality on social desirability scores was close to zero (Dodou & de Winter, 2014) and very small on measures of emotional functioning (Fouladi et al., 2002), future studies should aim to evaluate the cross-national measurement of the SPANE using the same administration mode across countries. It is also important to note that the data used in the present study were collected using an online mode of administration in 10 countries, whereas paper-and-pencil and telephone interviewing methods were used in two (India and Japan) and one country (China), respectively. The large sample disproportion hindered us from testing measurement invariance across the three administration modes, which is also an important avenue for future research. Fifth, to achieve acceptable model fit in three countries, we had to introduce correlated residuals between some items appearing near to each other within the SPANE (i.e., the first four items). The positioning of items is a potential limitation of the SPANE, because the first six items refer to general emotional experiences, with each negative emotion (viz., negative, bad, unpleasant) appearing after its conceptual opposite (viz., positive, good, pleasant), respectively. It would be useful to compare several versions of the SPANE with different item orders to evaluate possible item order effects (Rasinski et al., 2012).

Finally, the present study merely focused on examining the cross-national measurement invariance of the SPANE. Future research should also evaluate the convergent and discriminant validity of the scale across different cultures. Given the strong negative correlation between SPANE-P and SPANE-N found in all non-Asian countries, future cross-cultural studies should further investigate the independence and bipolarity of positive and negative emotions, as measured with the SPANE, by examining similarities and differences in predictors and outcomes of positive and negative emotions.

Conclusions

The SPANE can be recommended as a promising tool for cross-cultural studies of emotional experience and well-being, but further research on its measurement invariance across a more diverse set of countries is warranted. The evidence of partial scalar invariance indicates that the scale can be used with caution in research comparing correlates and mean levels of emotional experiences from a cross-cultural perspective. Our findings suggest that positive emotion

terms (both general and specific) and general negative emotion terms (e.g., negative, bad, unpleasant) display a greater level of invariance across languages than specific negative emotion terms, and thus might be more appropriate for use in cross-cultural studies on emotional experiences and well-being. To conclude, future cross-cultural research on the performance of the SPANE is warranted, but our findings offer initial evidence that this scale may be a valuable tool for measuring emotional experiences in adult samples from different countries.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The U.S. data collection was funded in part by the Marchionne Summer Research Fellowship from Washington State University and the Total Worker Health® Dissertation Award from the Oregon Healthy Workforce Center, a Total Worker Health® Center of Excellence funded by the National Institute for Occupational Safety and Health (Grant number U19OH010154).

ORCID iDs

Veljko Jovanović https://orcid.org/0000-0001-9248-2518
Marta Martín-Carbonell https://orcid.org/0000-0002-6337-577X
Begoña Espejo https://orcid.org/0000-0003-0788-2969
Julia Krasko https://orcid.org/0000-0002-8046-2330
Kwok Kit Tong https://orcid.org/0000-0001-8594-9300
Murat Yıldırım https://orcid.org/0000-0003-1089-1380

Supplemental Material

Supplemental material for this article is available online.

References

Asparouhov, T., & Muthen, B. (2010). *Weighted least squares estimation with missing data* (M*plus* Technical Appendix, pp. 1-10). http://www.statmodel.com/download/GstrucMissingRevision.pdf

Bagozzi, R. P., Wong, N., & Yi, Y. (1999). The role of culture and gender in the relationship between positive and negative affect. *Cognition and Emotion*, *13*(6), 641-672. https://doi.org/10.1080/026999399379023

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 186-203. https://doi.org/10.1207/s15328007sem1302_2

Busseri, M. (2018). Examining the structure of subjective well-being through meta-analysis of the associations among positive affect, negative affect, and life satisfaction. *Personality and Individual Differences*, *122*, 68-71. https://doi.org/10.1016/j.paid.2017.10.003

Byrne, B. M., Shavelson, R., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456-466. https://doi.org/10.1037/0033-2909.105.3.456

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Chen, F. F., Bai, L., Lee, J. M., & Jing, Y. (2016). Culture and the structure of affect: A bifactor modeling approach. *Journal of Happiness Studies*, *17*, 1801-1824. https://doi.org/10.1007/s10902-015-9671-3

Cieciuch, J., Davidov, E., Schmidt, P., & Algesheimer, R. (2019). How to obtain comparable measures for cross-national comparisons. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *71*, 157-186. https://doi.org/10.1007/s11577-019-00598-7

Cooke, P. J., Melchert, T. P., & Connor, K. (2016). Measuring well-being: A review of instruments. *The Counseling Psychologist*, *44*(5), 730-757. https://doi.org/10.1177/0011000016633507

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *43*(Pt. 3), 245-265. https://doi.org/10.1348/0144665031752934

Daniel-González, L., de la Rubia, J. M., Valle de la, O. A., García-Cadena, C. H., & Martínez-Martí, M. L. (2020). Validation of the Mexican Spanish version of the Scale of Positive and Negative Experience in a sample of medical and psychology students. *Psychological Reports*, *123*(5), 2053-2079. https://doi.org/10.1177/0033294119896046

Davis, R. C., Arce, M. A., Tobin, K. E., Palumbo, I. M., Chmielewski, M., Megreya, A. M., & Latzman, R. D. (2020). Testing measurement invariance of the Positive and Negative Affect Schedule (PANAS) in American and Arab university students. *International Journal of Mental Health and Addiction*. Advance online publication. https://doi.org/10.1007/s11469-020-00411-z

Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, *95*(3), 542-575. https://doi.org/10.1037/0033-2909.95.3.542

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment*, *49*(1), 71-75. https://doi.org/10.1207/s15327752jpa4901_13

Diener, E., Oishi, S., & Tay, L. (2018). Advances in subjective well-being research. *Nature Human Behaviour*, *2*, 253-260. https://doi.org/10.1038/s41562-018-0307-6

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, *125*(2), 276-302. https://doi.org/10.1037/0033-2909.125.2.276

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

Dodou, D., & de Winter, J. C. F. (2014). Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior*, *36*, 487-495. https://doi.org/10.1016/j.chb.2014.04.005

du Plessis, G. A., & Guse, T. (2017). Validation of the Scale of Positive and Negative Experience in a South African student sample. *South African Journal of Psychology*, *47*(2), 184-197. https://doi.org/10.1177/0081246316654328

Epstein, J., Santo, R. M., & Guillemin, F. (2015). A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *Journal of Clinical Epidemiology*, *68*(4), 435-441. https://doi.org/10.1016/j.jclinepi.2014.11.021

Espejo, B., Checa, I., Perales-Puchalt, J., & Lisón, J. F. (2020). Validation and measurement invariance of the Scale of Positive and Negative Experience (SPANE) in a Spanish general sample. *International Journal of Environmental Research and Public Health*, *17*(22), 1-15. https://doi.org/10.3390/ijerph17228359

Feldt, L. S., Woodru, D. J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement*, *11*(1), 93-103. https://doi.org/10.1177/014662168701100107

Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research*, *18*(3), 382-388. https://doi.org/10.2307/3150980

Fouladi, R. T., McCarthy, C. J., & Moller, N. (2002). Paper-and-pencil or online? Evaluating mode effects on measures of emotional functioning and attachment. *Assessment*, *9*(2), 204-215. https://doi.org/10.1177/10791102009002011

Grossmann, I., Huynh, A. C., & Ellsworth, P. C. (2016). Emotional complexity: Clarifying definitions and cultural correlates. *Journal of Personality and Social Psychology*, *111*(6), 895-916. https://doi.org/10.1037/pspp0000084

Giuntoli, L., Ceccarini, F., Sica, C., & Caudek, C. (2017). Validation of the Italian versions of the Flourishing Scale and of the Scale of Positive and Negative Experience. *SAGE Open*, *7*(1), 1-12. https://doi.org/10.1177/2158244016682293

Hair, J., Black, W., Babin, B., & Anderson, R. (2010). *Multivariate data analysis* (7th ed.). Prentice Hall, Inc.

Helliwell, J. F., Layard, R., Sachs, J., & De Neve, J-E. (2020). *World happiness report 2020*. Sustainable Development Solutions Network.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind* (3rd ed.). McGraw-Hill.

Jackson, P., & Agunwamba, C. (1977). Lower bounds for the reliability of the total score on a test composed of nonhomogeneous items: I: Algebraic lower bounds. *Psychometrika*, *42*, 567-578. https://doi.org/10.1007/BF02295979

Jackson, J. C., Watts, J., Henry, T. R., List, J.-M., Forkel, R., Mucha, P. J., Greenhill, S. J., Gray, R. D., & Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. *Science*, *366*(6472), 1517-1522. https://doi.org/10.1126/science.aaw8160

Jang, S., Kim, E. S., Cao, C., Allen, T. D., Cooper, C. L., Lapierre, L. M., O'Driscoll, M. P., Sanchez, J. I., Spector, P. E., Peolmans, S. A. Y., Abarca, N., Alexandrova, M., Antoniou, A.-S., Beham, B., Brough, P., Carikci, I., Ferreiro, P., Fraile, G., Geurts, S., . . .Woo, J-M. (2017). Measurement invariance of the Satisfaction With Life Scale across 26 countries. *Journal of Cross-Cultural Psychology*, *48*(4), 560-576. https://doi.org/10.1177/0022022117697844

Joshanloo, M., & Bakhshi, A. (2016). The factor structure and measurement invariance of positive and negative affect: A study in Iran and the USA. *European Journal of Psychological Assessment*, *32*(4), 265-272. https://doi.org/10.1027/1015-5759/a000252

Jovanović, V., & Brdar, I. (2018). The cross-national measurement invariance of the Satisfaction with Life Scale in a sample of undergraduate students. *Personality and Individual Differences*, *128*, 7-9. https://doi.org/10.1016/j.paid.2018.02.010

Jovanović, V., Lazić, M., Gavrilov-Jerković, V., & Molenaar, D. (2020). The Scale of Positive and Negative Experience (SPANE): Evaluation of measurement invariance and convergent and discriminant validity. *European Journal of Psychological Assessment*, *36*(4), 694-704. https://doi.org/10.1027/1015-5759/a000540

Kang, Y., McNeish, D. M., & Hancock, G. R. (2016). The role of measurement quality on practical guidelines for assessing measurement and structural invariance. *Educational and Psychological Measurement*, *76*(4), 533-561. https://doi.org/10.1177/0013164415603764

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Kööts-Ausmees, L., Realo, A., & Allik, J. (2013). The relationship between life satisfaction and emotional experience in 21 European countries. *Journal of Cross-Cultural Psychology*, *44*(2), 223-244. https://doi.org/10.1177/0022022112451054

Krasko, J., Gollan, N., & Koch, T. (2019). *Digital Well-Being* [Unpublished raw data].

Krasko, J., Intelisano, S., & Luhmann, M. (2020). *The complexity of happiness conceptualizations are associated with the success of well-being related intentions and behaviors in everyday life* [Manuscript in preparation].

Krasko, J., Intelisano, S., & Luhmann, M. (2021). When happiness is both joy and purpose: The complexity of lay definitions of happiness and well-being is related to actual well-being. *PsyArXiv*. https://doi.org/10.31234/osf.io/6c8fm

Kyriazos, T. A., Stalikas, A., Prassa, K., & Yotsidi, V. (2018). A 3-faced construct validation and a bifactor subjective well-being model using the Scale of Positive and Negative Experience, Greek version. *Psychology*, *9*(5), 1143-1175. https://doi.org/10.4236/psych.2018.95071

Lee, S. T., Hartanto, A., Yong, J., Koh, B., & Leung, A. K. (2020). Examining the cross-cultural validity of the positive affect and negative affect schedule between an Asian (Singaporean) sample and a Western (American) sample. *Asian Journal of Social Psychology*, *23*(1), 109-116. https://doi.org/10.1111/ajsp.12390

Legatum Institute. (2020). *The Legatum Prosperity Index 2020*. https://www.prosperity.com/download_file/view_inline/4193

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*, 936-949. https://doi.org/10.3758/s13428-015-0619-7

Li, F., Bai, X., & Wang, Y. (2013). The Scale of Positive and Negative Experience (SPANE): Psychometric properties and normative data in a large Chinese sample. *PLOS ONE*, *8*, Article e61137. https://doi.org/10.1371/journal.pone.0061137

Linton, M., Dieppe, P., & Medina-Lara, A. (2016). Review of 99 self-report measures for assessing well-being in adults: Exploring dimensions of well-being and developments over time. *BMJ Open*, *6*, Article e010641. https://doi.org/10.1136/bmjopen-2015-010641

Liu, J-D., You, R-H., Liu, H., & Chung, P-K. (2020). Chinese version of the international positive and negative affect schedule short form: Factor structure and measurement invariance. *Health and Quality of Life Outcomes*, *18*, Article 285. https://doi.org/10.1186/s12955-020-01526-6

Mair, P. (2018). *Modern psychometrics with R*. Springer International.

Martín-Carbonell, M., Checa, I., Fernández-Daza, M., Paternina, Y., & Espejo, B. (in press). Adaptation and psychometric properties of the Scale of Positive and Negative Experience (SPANE) in the general Colombian population. *International Journal of Environmental Research and Public Health*.

Martín-Carbonell, M., Espejo, B., Checa, I., & Fernández-Daza, M. (2021). Adaptation and measurement invariance by gender of the Flourishing Scale in a Colombian sample. *International Journal of Environmental Research and Public Health*, *18*(5), 2664. https://doi.org/10.3390/ijerph18052664

Martin-Krumm, C., Fenouillet, F., Csillik, A., Kern, L., Besancon, M., Heutte, J., Paquet, T., Delas, Y., Trousselard, M., Lecorre, B., & Diener, E. (2018). Changes in emotions from childhood to young adulthood. *Child Indicators Research*, *11*, 541-561. https://doi.org/10.1007/s12187-016-9440-9

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

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*(3), 568-592. https://doi.org/10.1037/0021-9010.93.3.568

Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. Routledge.

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, *39*(4), 479-515. https://doi.org/10.1207/S15327906MBR3903_4

Mitamura, C., Leu, J., Campos, B., Boccagno, C., & Tugade, M. M. (2014). Traversing affective boundaries: Examining cultural norms for positive emotions. In M. M. Tugade, M. N. Shiota, & L. D. Kirby (Eds.), *Handbook of positive emotions* (pp. 229-240). Guilford Press.

Miyamoto, Y., & Ryff, C. D. (2011). Cultural differences in the dialectical and non-dialectical emotional styles and their implications for health. *Cognition and Emotion*, *25*(1), 22-39. https://doi.org/10.1080/02699931003612114

Mroczek, D. K., & Kolarz, C. M. (1998). The effect of age on positive and negative affect: A developmental perspective on happiness. *Journal of Personality and Social Psychology*, *75*(5), 1333-1349. https://doi.org/10.1037/0022-3514.75.5.1333

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Oishi, S., Graham, J., Kesebir, S., & Galinha, I. C. (2013). Concepts of happiness across time and cultures. *Personality and Social Psychology Bulletin*, *39*(5), 559-577. https://doi.org/10.1177/0146167213480042

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, *46*, 1023-1031. https://doi.org/10.3758/s13428-013-0434-y

Putnick, D. L., & Bornstein, M. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71-90. https://doi.org/10.1016/j.dr.2016.06.004

Rahm, T., Heise, E., & Schuldt, M. (2017). Measuring the frequency of emotions-validation of the Scale of Positive and Negative Experience (SPANE) in Germany. *PLOS ONE*, *12*, Article e0171288. https://doi.org/10.1371/journal.pone.0171288

Rasinski, K. A., Lee, L., & Krishnamurty, P. (2012). Question order effects. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbooks in psychology®. APA handbook of research methods in psychology, Vol. 1. Foundations, planning, measures, and psychometrics* (pp. 229-248). American Psychological Association.

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

Rice, S. P. M. (2020a). *Measuring well-being: A longitudinal psychometric investigation into the theoretical structures of well-being and implications for assessment* (Publication No. 27739173) [Doctoral dissertation, Washington State University]. ProQuest Dissertations Publishing.

Rice, S. P. M. (2020b). *Psychometric properties of well-being scales in and outside the workplace* [Unpublished raw data]. Washington State University Vancouver.

Rice, S. P. M., & Shorey-Fennell, B. R. (2020). Comparing the psychometric properties of measures of positive and negative emotional experiences: Implications for the assessment of subjective wellbeing. *Journal of Well-Being Assessment*, *4*, 37-56. https://doi.org/10.1007/s41543-020-00025-1

Saucier, G., Thalmayer, A. G., & Bel-Bahar, T. S. (2014). Human attribute concepts: Relative ubiquity across twelve mutually isolated languages. *Journal of Personality and Social Psychology*, *107*(1), 199-216. https://doi.org/10.1037/a0036492

Scherer, K. R., Shuman, V., Fontaine, J. R. J., & Soriano, C. (2013). The GRID meets the wheel: Assessing emotional feeling via self-report. In J. R. J. Fontaine, K. R. Scherer, & C. Soriano (Eds.), *Components of emotional meaning: A sourcebook* (pp. 281-298). Oxford University Press.

Schimmack, U., Oishi, S., & Diener, E. (2002). Cultural influences on the relation between pleasant emotions and unpleasant emotions: Asian dialectic philosophies or individualism-collectivism? *Cognition and Emotion*, *16*(6), 705-719. https://doi.org/10.1080/02699930143000590

Scollon, C. N., Koh, S., & Au, E. W. M. (2011). Cultural differences in the subjective experience of emotion: When and why they occur. *Social and Personality Psychology Compass*, *5*(11), 853-864. https://doi.org/10.1111/j.1751-9004.2011.00391.x

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

Singh, K., Junnarkar, M., & Kaur, J. (2016). *Measures of positive psychology: Development and validation*. Springer.

Spencer-Rodgers, J., Peng, K., & Wang, L. (2010). Dialecticism and the co-occurrence of positive and negative emotions across cultures. *Journal of Cross-Cultural Psychology*, *41*(1), 109-115. https://doi.org/10.1177/0022022109349508

Sumi, K. (2014). Reliability and validity of Japanese versions of the Flourishing Scale and the Scale of Positive and Negative Experience. *Social Indicators Research*, *118*, 601-615. https://doi.org/10.1007/s11205-013-0432-6

Tabachnick, B. G., & Fidell, L. S. (2015). *Using multivariate statistics*. Allyn & Bacon.

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology*, *38*(2), 227-242. https://doi.org/10.1177/0022022106297301

Thompson, B., Roberts, S., & Lupyan, G. (2020). Cultural influences on word meanings revealed through large-scale semantic alignment. *Nature Human Behaviour*, *4*, 1029-1038. https://doi.org/10.1038/s41562-020-0924-8

Tong, K. K., & Wang, Y. Y. (2017). Validation of the flourishing scale and scale of positive and negative experience in a Chinese community sample. *PLOS ONE*, *12*, Article e0181616. https://doi.org/ 10.1371/journal.pone.0181616

United Nations Development Programme. (2020). *Human development report 2020. The next frontier: Human development and the Anthropocene*.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4-69. https://doi.org/10.1177/109442810031002

Wang, J., & Wang, X. (2019). *Structural equation modeling: Applications using Mplus*. John Wiley & Sons.

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. https://doi.org/10.1037//0022-3514.54.6.1063

Werts, C. E., Linn, R. N., & Joreskog, K. G. (1974). Interclass reliability estimates: Testing structural assumptions. *Educational and Psychological Measurement*, *34*(1), 25-33. https://doi.org/10.1177/001316447403400104

Wilken, B., & Miyamoto, Y. (2018). Dialectical emotions. In J. Spencer-Rodgers, & K. Peng (Eds.), *The psychological and cultural foundations of East Asian cognition: Contradiction, change, and holism* (pp. 509-546). Oxford University Press.

Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, *81*, 1014-1045. https://doi.org/10.1007/s11336-016-9506-0

Yik, M., Russell, J. A., & Steiger, J. H. (2011). A 12-point circumplex structure of core affect. *Emotion*, *11*(4), 705-731. https://doi.org/10.1037/a0023980

Yildirim, M. (2019). Mediating role of resilience in the relationships between fear of happiness and affect balance, satisfaction with life, and flourishing. *Europe's Journal of Psychology*, *15*(2), 183-198. https://doi.org/10.5964/ejop.v15i2.1640

Yoon, M., & Lai, M. H. C. (2018). Testing factorial invariance with unbalanced samples. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(2), 201-213. https://doi.org/10.1080/10705511.2017.1387859