Psychological Assessment

Testing Measurement Invariance of the Depression, Anxiety, and Stress Scales (DASS-21) Across Four Countries

Saskia Scholten, Julia Velten, Angela Bieda, Xiao Chi Zhang, and Jürgen Margraf

Online First Publication, January 26, 2017. http://dx.doi.org/10.1037/pas0000440

Psychological Assessment

1040-3590/17/$12.00 http://dx.doi.org/10.1037/pas0000440

Testing Measurement Invariance of the Depression, Anxiety, and Stress Scales (DASS-21) Across Four Countries

Saskia Scholten, Julia Velten, Angela Bieda, Xiao Chi Zhang, and Jürgen Margraf
Ruhr-Universität Bochum

The rising burden of mental and behavioral disorders has become a global challenge (Murray et al., 2012). Measurement invariant clinical instruments are necessary for the assessment of relevant symptoms across countries. The present study tested the measurement invariance of the 21-item version of the Depression, Anxiety, and Stress Scales (DASS; Lovibond & Lovibond, 1995b) in Poland, Russia, the United Kingdom (U.K.), and the United States of America (U.S.). Telephone interviews were conducted with population-based samples ($n_{PL} = 1003$, $n_{RU} = 3020$, $n_{U.K.} = 1002$, $n_{U.S.} = 1002$). The DASS-21 shows threshold measurement invariance. Comparisons of latent means did not indicate differences between U.K. and U.S. samples. However, Polish and Russian samples reported more depressive symptoms compared with U.K. and U.S. samples; the Russian sample had the highest levels of anxiety symptoms and the Polish sample demonstrated the highest stress levels. The DASS-21 can be recommended to meaningfully compare the relationships between variables across groups and to compare latent means in Polish-, Russian-, and English-speaking populations.

Public Significance Statement
This study underlines the importance to test equivalence and examine potential bias in cross-cultural research frameworks. We successfully tested the cross-cultural applicability of a questionnaire for symptoms of depression, anxiety, and stress and found that the populations in Poland and Russia reported more symptoms of depression, anxiety, and stress than the populations in the U.K. and the U.S.

Keywords: cross-cultural comparison, Depression, Anxiety and Stress Scales, measurement invariance, psychological assessment

Supplemental materials: http://dx.doi.org/10.1037/pas0000440.supp

Across countries the increase of anxiety and mood disorders has become a widespread challenge. Worldwide, anxiety disorders are found most often with a 12-month prevalence of 11%. Although the average 12-month prevalence of mood disorders is slightly lower (6%), mood disorders are still the second largest group of mental disorders worldwide (Kessler et al., 2009). The global impact of depressive and anxiety disorders becomes especially salient in terms of disability adjusted life years (DALY). Depressive disorders accounted for 41% and anxiety disorders accounted for 15% of DALY caused by mental and substance use disorders in 2010 (Murray et al., 2012).

However, prevalence rates vary worldwide: In Europe, 38% of the population suffers from a mental disorder each year. Anxiety disorders take up the greatest part with 14%. Major depression accounts for 7% of the disorders (Wittchen et al., 2011). In the United States the picture is similar; 12-month prevalence of anxiety disorders is 18%, followed by mood disorders with 10% (Kessler, Chiu, Demler, & Walters, 2005). But in Central Asia estimated point prevalence for anxiety disorders is 4% for men and 6% in women (Baxter et al., 2014) whereas symptoms of anxiety were much higher in a population-based survey in Arkhangelsk in Russia (22% of the men and 53% of the women), again followed by symptoms of depression (11% of the men and 34% of the women) (Averina et al., 2005). In a cross-sectional study of urban population samples in Russia, Poland, and the Czech Republic, the point-prevalence of depressive symptoms was around 20% in men and 40% in women (Bobak et al., 2006).

Differences in prevalence rates do not necessarily represent real cross-cultural differences. Specific symptoms might belong to the

Saskia Scholten, Julia Velten, Angela Bieda, Xiao Chi Zhang, and Jürgen Margraf, Faculty of Psychology, Department of Clinical Psychology and Psychotherapy, Mental Health Research and Treatment Center, Ruhr-Universität Bochum.

Support by the Alexander von Humboldt Professorship awarded to Jürgen Margraf by the Alexander von Humboldt Foundation is gratefully acknowledged. We also thank the authors of the Polish version of the DASS-21 for the translation as well as Gerrit Hirschfeld, Sunthud Pornprasertmanit, Yves Rosseel and Manfred Schmitt for methodological consultations. Finally, we would also like to thank Helen Vollrath for her English proof.

Correspondence concerning this article should be addressed to Saskia Scholten, Faculty of Psychology, Department of Clinical Psychology and Psychotherapy, Mental Health Research and Treatment Center, Ruhr-Universität Bochum, Massenbergstr. 9-13, 44787 Bochum, Germany. E-mail: saskia.scholten@rub.de

range of culturally normative experiences in one culture and might be actual signs of distress in another (Kirmayer, 2001; López & Guarnaccia, 2000). So called culture-bound syndromes describe a cluster of symptoms that only appear in a specific cultural context (Weiss & Somma, 2007). Yet, even if the same symptom cluster appears across cultures, methodological issues can still lead to false conclusions about cross-cultural differences (Bowden & Fox-Rushby, 2003; van de Vijver & Leung, 1997; van de Vijver & Tanzer, 2004). For example, the description of a certain disorder might differ across countries, making a word-by-word translation unfeasible to assess symptoms of the specific disorder (Ryder et al., 2008); or systematic bias in the assessment itself might appear as a result of answer tendencies or familiarity with the question format (Byrne, 2016). More than 50 years of research have addressed these problems (Poortinga, 1995), but best-practice strategies of cross-cultural assessment and comparisons are just recently entering mainstream psychological research (Byrne, 2016).

The purpose of the present study was to test the cross-cultural applicability of the 21-item version of the Depression, Anxiety, and Stress Scales (DASS-21; Lovibond & Lovibond, 1995a) in two Western and two Eastern European countries[1] and to compare the means of symptoms of depression, anxiety, and stress across those countries if applicability is given. In the introduction, we therefore first outline central issues of cross-cultural assessment. Then, we provide an overview of the DASS-21 and its applicability within and across cultures today. Finally, we describe the aim of the study in detail.

Cross-Cultural Assessment

The earliest approach to cross-cultural psychology was *transport and test* (Berry, 1969; Berry, Poortinga, Segall, & Dasen, 2002). Here, tests and interventions established in Europe or the U.S. were literally exported and imposed on local cultures without cultural adaptation (F. M. Cheung, 2012). Without further reflection, the Euro/American-centered view was assumed to be universal. Today, cross-cultural differences are taken into account as important sources of information to understand psychosocial mechanisms of human behavior, interaction, and society. To avoid former culture-blind mistakes, it is important that research approaches are equally valid across cultures. van de Vijver and Leung (1997) distinguish between three levels of equivalence: construct equivalence, measurement equivalence, and scalar equivalence. With respect to construct equivalence, the same constructs are under study across different cultural groups even though they might be assessed differently. If metric measures have the same measurement unit, measurement equivalence is given. Finally, scalar equivalence describes a case where not only the measurement unit is the same, but also the origin or scale.

Within cross-cultural psychology, special consideration needs to be given to cross-cultural assessment. Still too often, it is assumed that a translated instrument measures the same phenomena in the same way in a different culture (Byrne, 2016). However, this assumption needs to be tested and if it does not hold, the measurement instrument should be adapted accordingly. Most frequently assessment instruments are only *adopted* from one culture to another which refers to a close translation of an instrument into a target language (He & van de Vijver, 2012). The procedure is valid if the construct and the instrument features are suitable in all cultural groups involved. Yet, this condition is rarely fulfilled. *Adaptation* of assessment instruments considers this problem and expands the mere translation as a linguistic task through psychometric evaluation of the instrument in the new context (He & van de Vijver, 2012). An example of a detailed description of test adaptation procedures is provided in the guidelines of the International Test Commission (Hambleton, 2005; International Test Commission, 2005, 2016). If adoption and adaptation do not yield satisfactory results, a new instrument needs to be compiled which is called *assembly*. The new instrument should be maximally culturally suitable, however, throughout the process the comparability across cultures decreases (He & van de Vijver, 2012).

For the present study, cross-cultural adaptation is the relevant approach because we aimed to test the cross-cultural applicability of the DASS-21. First, both the instrument and the measured construct need to be rigorously validated for the conversion of an assessment scale from one language and cultural application for the use in another language and culture. Then, the levels of equivalence mentioned above need to be tested for the respective assessment instrument (Byrne, 2016). In this more specific context, the term *measurement invariance* is used synonymously for measurement equivalence to describe equivalence as a property of a measurement instrument (Davidov, Meuleman, Cieciuch, Schmidt, & Billiet, 2014). To evaluate whether a translated version is comparable across countries, commonly three successive more restricted levels of invariance or equivalence are tested (Byrne, 2008, 2016; Hirschfeld & Von Brachel, 2014; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000): On the first level *configural invariance* is tested. It implies that the configuration of the model, the number of factors and their loading patterns, are the same across groups without imposing further equality constraints on the model parameters. On the second level, all factor loadings are constrained to be equal across groups to test *metric invariance*. It tests to what extent the items designed to measure the same factor are equivalent across groups. If the assumption holds, the factorial validity of an instrument is equivalent across samples. In contrast, latent means can only be meaningfully compared across different groups if, in addition to the factor loadings, also item intercepts are equal across groups. *Scalar invariance* implies that not only factor loadings, but also item intercepts are operating equivalently across groups. In models with ordered-categorical variables, instead of intercepts, thresholds are restricted testing for *threshold invariance.*

If this testing procedure reveals noninvariant items either on a metric or a scalar/threshold level, it cannot be assumed that the items of the translated instrument measure the same factor equally across groups. Occurrence of noninvariant item parameters relate to biases that can become evident on any level of equivalence and stem from differences in the measured constructs (construct bias), methodological artifacts such as familiarity with the item format (measurement bias) or problematic item content leading to cross-cultural differences in the interpretation of the item (item bias) (Borsboom, 2006; Byrne, 2016; van de Vijver & Tanzer, 2004).

[1] The authors are aware that country boundaries do not necessarily reflect homogenous cultural groups. However, the study described in this article has been conducted on a national level. Therefore, country and culture are both used synonymously to describe the groups underlying this study.

Byrne, Shavelson, and Muthén (1989) introduced the approach of *partial measurement invariance* as an approach to treat noninvariant parameters. Here, some parameters are still constrained equal across groups whereas noninvariant parameters are identified and estimated freely (Byrne, 2012; Chen, 2008; Millsap & Kwok, 2004; Sass, 2011; Steinmetz, 2013). Alternatively, *assembly* can be considered as a strategy to deal with noninvariance.

Even though the issue of measurement invariance has been discussed in cross-cultural research for some decades (van de Vijver & Leung, 1997; Vandenberg & Lance, 2000), measurement invariant instruments in cross-cultural research within clinical psychology are still an exception (Byrne, Stewart, Kennard, & Lee, 2007; Crockett, Randall, Shen, Russell, & Driscoll, 2005; Nuevo et al., 2009; Torres, Miller, & Moore, 2013). Examples are the Brief Symptom Inventory (Hoe & Brekke, 2009), the Patient Health Inventory (Ryan, Bailey, Fearon, & King, 2013), the General Behavior Questionnaire (Pendergast et al., 2015), or Beck's Depression Inventory (Byrne et al., 2007; Canel-Çınarbaş, Cui, & Lauridsen, 2011; Dere et al., 2015; Nuevo et al., 2009; Whisman, Judd, Whiteford, & Gelhorn, 2013). Testing measurement invariance of the DASS-21 across four countries, the current research contributes to the cross-cultural applicability of this instrument and extends the small number of comparable cross-cultural assessment instruments for symptoms of the most widespread mental health problems.

Depression, Anxiety, and Stress Scales

Based on the tripartite model of anxiety and depression (Clark & Watson, 1991), Lovibond and Lovibond (1995a) developed the Depression, Anxiety, and Stress Scales (DASS) that consists of 42 items and three dimensions (DASS-42). The depression scale comprises symptoms of anhedonia and inactivity (e.g., "I couldn't experience any positive feeling at all."). The anxiety scale describes symptoms of physiological hyperarousal and specific anxiety (e.g., "I found myself in situations that made me so anxious I was most relieved when they ended."). General distress as a common symptom of depression and anxiety disorders is conceptualized in the stress scale (e.g., "I found it difficult to relax."). A 21-item short version of the DASS with comparable psychometric properties was proposed by Lovibond and Lovibond (1995a) and subsequently validated (DASS-21; Antony, Bieling, Cox, Enns, & Swinson, 1998; Crawford & Henry, 2003). The short version rates depressive, anxiety, and stress symptoms over the past week on three 7-item subscales using 4-point Likert scales from 0 (*did not apply to me at all*) to 3 (*applied to me very much or most of the time*).

The DASS have several practical advantages: First, they are brief and can be applied in clinical and nonclinical samples (Antony et al., 1998; Brown, Chorpita, Korotitsch, & Barlow, 1997; Gloster et al., 2008; Gomez, Summers, Summers, Wolf, & Summers, 2014; Page, Hooke, & Morrison, 2007). In clinical samples, the DASS-42 distinguishes between anxiety and mood disorder patients (Brown et al., 1997) and assesses symptoms of depression and anxiety comorbid to other disorders (Smith, Sullivan, Hopkins, & Douglas, 2004; Szabó, 2011; Taylor, Lovibond, Nicholas, Cayley, & Wilson, 2005; Wilson & Rapee, 2005). Second, the DASS-21 can be used as a routine clinical outcome measure (Ng et al., 2007). Besides, it can serve to classify whether a patient moved into the range of normal functioning (Ronk, Korman, Hooke, & Page, 2013). Finally, the DASS-21 and the DASS-42 can serve as screening instruments (Dahm, Wong, & Ponsford, 2013; Kok, de Haan, van der Meer, Najavits, & De Jong, 2015; Nieuwenhuijsen, de Boer, Verbeek, Blonk, & van Dijk, 2003). Bayram and Bilgel (2008) propose the DASS-42 as a screening instrument to implement adequate primary and secondary prevention measures. For English-speaking populations percentile norms from an Australian general adult population sample (Crawford, Cayley, Lovibond, Wilson, & Hartley, 2011) as well as for a large psychiatric outpatient sample (Davies, Caputi, Skarvelis, & Ronan, 2015) exist for both versions of the DASS. It allows researchers and practitioners to compare the individual score in reference to the general population for screening and monitoring purposes. It must be kept in mind that most of these studies were conducted in English-speaking countries and thus the percentile norms do not necessarily apply internationally. Despite this limitation, all of the practical implications delineated above are encouraging to promote the implementation of the DASS-21 across countries. As a precondition, the challenges of cross-cultural adaptation need to be considered; namely, the scales' translation, within-country validation and testing for measurement invariance across countries.

The DASS have been translated into 42 languages (Lovibond, 2015), which makes the scales widely accessible for practitioners and researchers. Yet, studies examining psychometric properties for the translated questionnaires are rare. Implicitly, researchers seem to assume that psychometric properties of the original questionnaire are transferrable and comparable for the translated version (Borsboom, 2006). However, this might not always be the case. Psychometric properties need to be investigated within the new population to ensure that the translated version functions like the original instrument. A literature search on PubMed, ISI, Web of Science, and PsychInfo in June 2016 revealed 14 published validations of the questionnaire excluding the original English version. Psychometric properties were reported for the DASS-42 in Brazilian Portuguese (Sardá, Nicholas, Pimenta, & Asghari, 2008), Chinese (Chan et al., 2012), Dutch (Nieuwenhuijsen et al., 2003), Italian (Severino & Haynes, 2010), and Turkish (Hekimoglu et al., 2012; Uncu, Bayram, & Bilgel, 2007). Validated and published versions of the DASS-21 exist in Brazilian Portuguese (Vignola & Tucci, 2014), Chinese (Gong, Xie, Xu, & Luo, 2010; Wang et al., 2016), Farsi (Asghari, Saed, & Dibajnia, 2008; Bayani, 2010; Sahebi, Asghari, & Salari, 2005), Indian (Singh, Junnarkar, & Sharma, 2015), Italian (Bottesi et al., 2015), German (Nilges & Essau, 2015), Malaysian (Musa, Fadzil, & Zain, 2007; Ramli, Ramli, Kartini, & Rosnani, 2011; Ramli, Salmiah, & Nurul Ain, 2009), Nepali (Tonsing, 2014), Portuguese (Apóstolo, Mendes, & Azeredo, 2006; Vasconcelos-Raposo, Fernandes, & Teixeira, 2013), Serbian (Jovanovic, Gavrilov-Jerkovic, Zuljevic, & Brdaric, 2014), Spanish (Antúnez & Vinet, 2012; Bados, Solanas, & Andrés, 2005; Daza, Novy, Stanley, & Averill, 2002), and Vietnamese (Tran, Tran, & Fisher, 2013). Almost all of the studies included values for internal reliability reported as Cronbach's alpha. Principal component analysis or confirmatory factor analysis as well as correlations with other psychometric measures were investigated in the majority of the studies. Only a small number of studies reported comparisons between clinical and nonclinical samples.

The DASS-21 has been tested only twice across different national groups. A study that tested measurement invariance across Asian cultures, namely Indonesia, Malaysia, Singapore, Sri Lanka, Taiwan, and Thailand had to revise the stress scale as it did not seem appropriate to Asian cultures (Oei, Sawang, Goh, & Mukhtar, 2013). Even though the three-factor model fitted the data well, the DASS-21 was not invariant across the Asian countries. Results of the second study suggested strong measurement invariance for the DASS-21 in paired comparisons of Australian and Chinese, as well as Australian and Malaysian, and Australian and Chilean adolescent samples (Mellor et al., 2015). Therefore, it is evident that measurement invariance testing of the DASS-21 should be continued to enhance the validity and the applicability of the DASS-21 internationally. Up till now, measurement invariance has not been tested in Eastern European or Central Asian countries such as Poland and Russia. This study aims to close this research gap.

Aim of the Study

The present study seeks to test measurement invariance of the DASS-21 in Poland and Russia as two Eastern European countries in reference to the United Kingdom (U.K.) and the United States of America (U.S.) as two representative Western countries. Additionally, latent mean differences are compared for symptoms of depression, anxiety, and stress provided that the DASS-21 is scalar measurement invariant. The study adds to the existing evidence of the cross-cultural applicability of the DASS-21 and enhances the options for practical implementation and cross-cultural research in the respective countries.

Method

Procedure

The Ethical Committee of the Faculty of Psychology at the Ruhr-Universität Bochum in Germany formally approved the study. Population-based surveys were conducted in Poland, Russia, the U.K., and the U.S. The study aimed to assess a representative sample of the population in all of the respective countries. To accomplish representativeness, each sample was drawn from the residential population aged 18 years and above that was accessible via landline or mobile phones. Landline telephone numbers were chosen based on regional stratification while mobile phone numbers were stratified by providers.

An independent social market and research institute conducted the survey. In Russia, the survey was conducted from December 2013 to February 2014 and resulted in a total number of 3,022 completed interviews. Two interviews had to be excluded because the denoted age did not match the year of birth stated. 3,007 interviews were conducted in Poland, the U.K., and the U.S. from June to October 2014. Kish selection grid was implemented to identify the member within a household to be interviewed. Prior to each interview, the interviewer introduced the study and gathered participants' informed consent. Response rate estimates of the proportion of actual eligible cases out of cases of unknown eligibility by the American Association for Public Opinion Research standards (The American Association for Public Opinion Research, 2016) varied between 13.4% and 20.1%.

Participant Characteristics

In Russia, the sample comprised 3,020 participants. The Polish, the U.K., and the U.S. samples approximately constituted 1,000 participants per country. This should be Table 1. shows the demographic characteristics of the samples in each country. Whereas variations in mean age and retirement status might represent differences of the populations, the disparity of the highest level of education is partly caused by slightly different methodological assessment of the educational level across countries.

Measures

Demographic variables. Demographic variables were gathered according to internationally recognized standards. Gender, age, marital status, and current employment status were assessed according to the recommendations of the Eurostat task force on social variables (Eurostat, 2007). Highest educational level was assessed differently across countries corresponding to the respective educational system (European Social Survey, 2012). Harmonization of educational levels was guided by the International Standard Classification of Education (ISCED; UNESCO Institute for Statistics, 2013). Because of the differing operationalization, the Russian educational levels did not perfectly match the ISCE-Classification and could be only categorized approximately.

Depression, Anxiety, and Stress Scales. The freely available English version of the DASS-21 was used in the United Kingdom and the United States (Antony et al., 1998; Henry & Crawford, 2005; Lovibond & Lovibond, 1995b). The Polish version of the DASS-21 was retrieved from the DASS website (Lovibond, 2015). For the purpose of another study, our team had developed a Russian version of the DASS-21 that was included in the Russian survey. The translation procedure followed principles of good practice proposed for patient-related outcome measures by Wild et al. (2005). A native Russian speaking person with fluency in German managed the translation process. The project manager as well as a Russian translator carried out forward translation. An independent native psychologist who was not involved in the forward translation process reconciled both versions in collaboration with the translators. Back translation was carried out by two independent German speakers who were second generation Russian immigrants. Back translations were reviewed and adjusted by an expert panel consisting of the project manager and a German as well as a Russian clinical psychology professor.

Data Analysis

Statistical analyses were conducted using SPSS version 21 (IBM Corporation, 2011), Mplus (Muthén & Muthén, 2015), and R (R Development Core Team, 2013). For the analysis in R, the statistical packages lavaan (Rosseel, 2012) and psych (Revelle, 2015) were applied.

Descriptive statistics. SPSS missing values analysis indicated missing values at random. Missing values did not exceed 5% for all sociodemographic values except for highest educational level in

Table 1
Sociodemographic Variables of the Participants

Variable	PL	RU	UK	US
Participants, *n*	1003	3020	1002	1002
Gender, % female	47.9	53.2	53.6	52.5
Age, *M* (*SD*)	44.6 (16.1)	43.2 (17.1)	58.2 (17.8)	51.2 (16.9)
Marital status[c]				
Unmarried	24.4	27.1	22.2	21.8
Married/Legal partnership	61.9	53.2	46.8	53.8
Widowed	6.2	8.4	16.5	9.3
Divorced/Separated	7.4	9.7	14.3	13.7
Missing	.1	1.6	.3	1.5
Highest level of education[c]				
Primary education or below	18.8	.5	6.7	.5
Lower secondary education	27.5	3.2	8.9	2.9
Upper and post-secondary education	16.6	59.0	36.2	53.1
Tertiary education (e.g. BA[a], MA[b])	35.8	37.3	29.2	38.3
Doctoral degree	.9	—	3.8	2.4
Missing	.8	—	15.3	2.8
Current employment status[c]				
Paid work	60.3	52.2	39.5	48.9
Education	7.0	6.7	2.1	3.3
Unemployed	6.4	7.9	3.3	4.4
Permanently sick or disabled	.7	—	3.5	7.3
Retired	20.8	23.9	44.8	23.0
Housework, looking after children, other	4.2	8.0	4.1	6.0
Other	—	—	—	4.4
Missing	.6	1.2	2.7	2.8

Note. PL = Poland; RU = Russia; UK = United Kingdom; US = United States of America.
[a] BA = Bachelor. [b] MA = Master. [c] Frequencies are presented in percent.

the U.K. (15.3% missing values). All of the DASS-21 variables taken together, 1.5% of the values in Poland were missing, as were 9.2% in Russia, 3.5% in the U.K., and 5.1% in the U.S. All available information was used to estimate confirmatory factor analyses. The procedure was guided by Asparouhov and Muthén's (2010) proposal for missing data with weighted least squares estimation which is based on pairwise deletion.

Descriptive statistics were calculated with R. Skewness and kurtosis were calculated for each sample separately. An absolute value larger than 2 for skewness or larger than 7 for kurtosis was considered as reference for substantial non-normality as is recommended for samples larger than 300 (Kim, 2013; West, Finch, & Curran, 1995). Besides, Mardia's normalized estimate of multivariate kurtosis was calculated as a measure to test the normality assumption (Mardia, 1970). Cronbach's alpha indicated internal reliability and was considered acceptable above $\alpha \geq .70$ (Cronbach, 1951).

Testing measurement invariance. Measurement invariance of the DASS-21 across four countries was tested based on Jöreskog's (1971) multigroup confirmatory factor analysis using Mplus (Muthén & Muthén, 2015). The approach was specified for ordered-categorical variables by Millsap and Yun-Tein (2004), Muthén and Asparouhov (2002), and Muthén and Christoffersson (1981). The ordinal nature of our data was taken into account following their recommendations: Consistent with Norton (2007) and Mellor et al. (2015), we estimated parameters using robust weighted-least-squares (WLSMV; Flora & Curran, 2004; B. O. Muthén, 1984). WLSMV is recommended to estimate thresholds if fewer than five response categories are given (Beauducel & Herzberg, 2006). Delta parameterization was used because it is also recommended for ordered-categorical data (Muthén & Muthén, 2012).

Single-group confirmatory factor analyses. As baseline model, we tested the hypothesized three-factor model of the DASS-21 in all countries separately (Crawford et al., 2011; Daza et al., 2002; Lovibond & Lovibond, 1995a; Musa, Fadzil, & Zain, 2007; Sinclair et al., 2012). For this reason, single-group confirmatory factor analysis (CFA) was conducted for each country separately (for more detail see: Byrne, 2008; Hirschfeld & Von Brachel, 2014). For purposes of statistical identification, the first observed variable of each latent scale was used as indicator variable and fixed to 1. The baseline model is depicted in Figure 1.

Multigroup confirmatory factor analyses to test measurement invariance. For multigroup confirmatory factor analysis, we used the more commonly applied forward approach which means including sequentially additional model constraints (Bovaird & Koziol, 2012; Hirschfeld & Von Brachel, 2014; Sass, 2011). To include all possible model comparisons, we tested Poland as the first reference country compared with all other three countries, followed by Russia as reference country in comparison to the U.K. and the U.S., as well as the U.K. as reference country compared with the U.S. This procedure was implemented in all subsequent steps of the measurement invariance testing.

Testing configural invariance. As depicted in the introduction, *configural invariance* was tested first. This means that the baseline models that were established in each country separately were combined to form the multigroup model and tested simultaneously (Byrne, 2008). The three factors and their loading patterns are thus the same across groups, but not yet constrained to be equal across

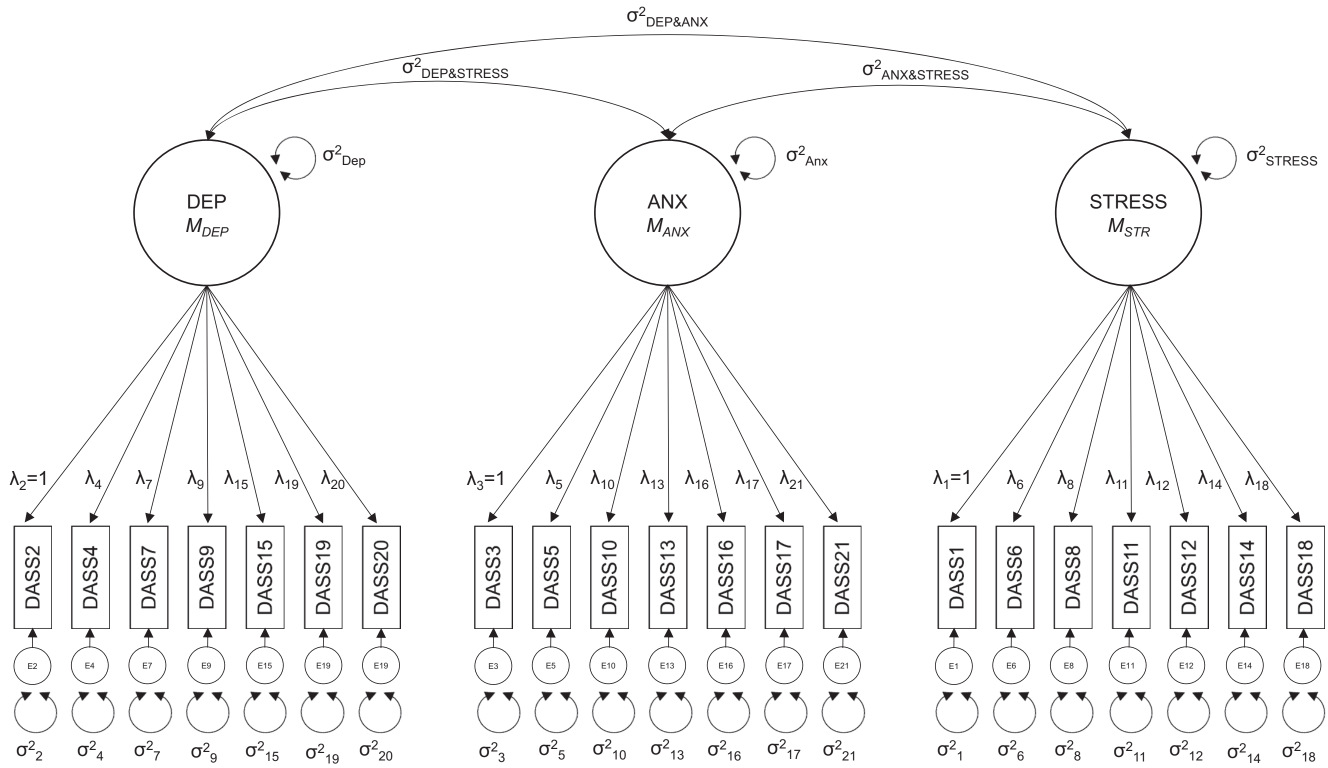

Figure 1. Baseline model tested for each country.

groups. To identify the multigroup model, factor means were fixed to 0 and scaling factors were fixed to 1 in all groups.

Testing metric invariance. Second, we tested *metric invariance* and constrained factor loadings to be equal across countries. More specifically, the factor loadings of the group that served as the reference group were estimated freely, whereas the factor loadings of the second, third, and fourth group were fixed to equal the factor loadings of the reference group (Byrne, 2008). The metric model was identified constraining the factor means of the reference group to 0 and constraining the scaling factors of the reference group to 1. The factor means and scaling factors of the other groups were estimated freely. Additionally, the first threshold of all variables as well as the second threshold of all indicator variables were constrained to be equal across groups (Millsap & Yun-Tein, 2004).

Testing threshold invariance. Finally, because of the ordinal structure of the present data *threshold invariance* was tested constraining all thresholds to be equal across countries. This model does not require additional restrictions for identification; only the factor mean was fixed to 0 and scaling factors were fixed to 1 in the reference group. If *threshold invariance* is met, *scalar invariance* can be assumed (Millsap & Yun-Tein, 2004).

Testing latent mean differences. Latent means can be compared after establishing threshold invariance (Byrne, 2012). The comparison was based on the model used to test threshold invariance. The latent means of one group were fixed to zero. This group served as the reference group for the comparison with the latent means of the other groups (Byrne, 2012). Hence, only latent mean differences can be interpreted. z scores and p values of the standardized model results are reported (L. K. Muthén & Muthén, 2012). Cohen's d was calculated as the effect size measure (small effect: $d \geq .20$, medium effect: $d \geq .50$, large effect: $d \geq .80$, Cohen, 1988).

Evaluating model fit. Typically, the Likelihood Ratio Test statistic that is expressed as chi-square (χ^2) statistic is used to evaluate the discrepancy between the covariance matrix of the restricted model and the covariance matrix of the unrestricted model (Byrne, 2012). A major limitation of using χ^2 is its sensitivity to large sample sizes which leads to oversized rejection rates (Byrne, 2012; Hirschfeld & Von Brachel, 2014). Therefore, we reported χ^2 only for reasons of sufficiency. A combination of two goodness-of-fit indices was used to evaluate our model: Root Mean Squared Error of Approximation (RMSEA; Steiger & Lind, 1980) and Comparative Fit Index (CFI; Bentler, 1990). All indices were based on robust values corrected in accordance with the WLSMV estimator. A CFI value of 1 would indicate perfect model fit. On the contrary, the value of 0 indicates perfect model fit for RMSEA. To ensure good model fit across different estimation methods, both goodness-of-fit indices were taken into account for the model evaluation (Schmitt, 2011). We followed Hu and Bentler's (1999) recommendations to evaluate model fit. According to them RMSEA $< .06–.08$ and CFI $> .95$ are cut-off criteria for a good model fit.

The model fit of a nested model was compared with a less restricted model with a scaled difference χ^2 test statistic (Asparouhov & Muthén, 2006; Satorra & Bentler, 2001). But as described above χ^2 is very sensitive to large sample sizes, therefore, χ^2 difference values are almost always significant. This problem prompted researchers (Chen, 2007; Cheung & Rensvold, 2002) to conduct Monte Carlo studies as a means to propose recommendations based on CFI and RMSEA difference values. ΔCFI and ΔRMSEA were obtained calculating the difference between the CFI values, or RMSEA values respectively, for the restricted model against the less restricted model. Therefore, we considered the drop in the CFI-value and the increase of the RMSEA-value to evaluate the change of model fit. A change of $\Delta CFI \geq .010$ accompanied by a change of $\Delta RMSEA \geq .015$ indicated a significant decrease in model fit and hence noninvariance (Chen, 2007).

Results

Descriptive Statistics

Boxplots presented as Supplemental Material 1 showed low symptom levels in all countries with the highest ratings on the stress scale. However, a number of outliers were found for all countries and for all scales. Table 2 shows the descriptive results by nation. According to Kim (2013) and West et al. (1995), skew and kurtosis indicated a normal distribution. However, Mardia's multivariate estimate for multivariate skew, Depression Scale: Median = 2.20, Anxiety Scale: Median = 3.05, Stress Scale = .75, and kurtosis, Depression Scale: Median = 5.01, Anxiety Scale: Median = 6.07, Stress Scale = 3.23, indicated a significant deviation from normality for all DASS-21 items. Internal consistency was ≥ .80 for all subscales except for the anxiety subscale in Poland, $\alpha = .78$. In all countries α was ≥ .90 for the total scale. All item correlations were significantly positive and ranged from $r = .14$, $p \leq .001$, U.K. sample: Item 1 and Item 7, to $r = .67$, $p \leq .001$, U.S. sample: Item 17 and Item 21. Supplemental Material 2 displays the detailed correlation tables.

Table 2
Descriptive Properties of the DASS-21 for Each Subscale Within Each Country

Country	n	Skew	Kurtosis	Mardia's skew	Mardia's kurtosis	α
Depression Scale						
PL	995	1.3	1.16	1.69***	4.16***	.84
RU	2820	1.36	1.83	1.84***	4.83***	.81
UK	982	1.78	3.06	3.17***	6.06***	.88
US	978	1.60	2.18	2.56***	5.18***	.88
Anxiety Scale						
PL	999	1.59	2.45	2.52***	5.45***	.78
RU	2934	1.68	3.02	2.83***	6.02***	.82
UK	990	1.99	4.50	3.96***	7.5***	.80
US	978	1.81	3.12	3.27***	6.12***	.84
Stress Scale						
PL	997	.66	−.30	.43***	2.7	.83
RU	2936	.95	.53	.9***	3.53***	.86
UK	987	.91	.37	.83***	3.37*	.86
US	977	.81	.09	.66***	3.09	.87
Total Scale						
PL	988	1.12	.82	1.25***	3.82***	.92
RU	2743	1.35	1.90	1.83***	4.9***	.93
UK	967	1.54	2.47	2.37***	5.47***	.93
US	951	1.40	1.70	1.96***	4.7***	.94

Note. PL = Poland; RU = Russia; UK = United Kingdom; US = United States of America.
* $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$.

Single-Group Confirmatory Factor Analysis

The results of the confirmatory factor analysis of the assumed three-factor model in each of the four countries suggested that the model was appropriate across all countries. Because the baseline model was well-fitting in all countries, model respecifications were not necessary. The model fit was shown line by line for each country in the upper part of Table 3 (for factor loadings of each scale by country, see Table 4). All CFI values exceeded the proposed cut-off score of > .95 for a good model fit (Hu & Bentler, 1999). The model fit was also good according to RMSEA, being smaller than the suggested cut-off score in all samples (Hu & Bentler, 1999).

Multi-Group Confirmatory Factor Analysis

Polish versus Russian, U.K., and U.S. samples. Results of the measurement invariance testing sequences are depicted in the lower part of Table 3 for each group comparison separately. In the first multigroup model, the Russian, the U.K., and the U.S. samples were compared with the Polish sample that was set as the reference country. Within this comparison, the first line shows the test of configural invariance. All fit indices indicated an appropriate model fit. The DASS-21 was also metric invariant as CFI was above .95 and RMSEA was below .08 for the metric model. ΔCFI was < .010 and ΔRMSEA < .015, which indicated that the more restricted model was as good as the less restricted model (Chen, 2007). When testing for threshold invariance, fit indices were still good. ΔCFI and ΔRMSEA did not indicate a meaningful drop of the model fit, $\Delta CFI = .003$, $\Delta RMSEA = .006$, (Chen, 2007). Thus, the Polish DASS-21 was threshold measurement invariant in comparison to the Russian, U.K., and U.S. questionnaire.

Russian versus U.K. and U.S. samples. The second multigroup model compared the U.K. and the U.S. to Russia. Again, fit indices of the configural model indicated configural invariance. Also, the metric model fitted the data well according to the fit indices and taking into account ΔCFI and ΔRMSEA, the model fit was as good as the configural model. The third line presents the results for the threshold model indicating a good model fit. Neither ΔCFI nor ΔRMSEA exceeded .01 and thus the model fit can be considered equal to the metric model (Chen, 2007). In sum, threshold invariance can be assumed for the Russian version in comparison to the U.K. and the U.S. versions of the DASS-21.

U.K. versus U.S. samples. Finally, the U.K. sample was compared with the U.S. sample in the third multigroup model. The configural model fitted the data well and so did both the metric and the threshold models. ΔCFI and ΔRMSEA dropped no more than .01 neither between the configural and the metric nor between the metric and the threshold model, thus threshold invariance can be assumed in this comparison as well.

Latent Mean Comparisons

Polish versus Russian, U.K., and U.S. samples. Figure 2 shows the latent mean differences with the Polish sample's latent mean fixed to zero serving as the reference country. Comparisons

Table 3
Fit-Indices for Single- and Multi-Group Confirmatory Factor Analyses

CFA	$\chi^2(df)^a$	CFI	RMSEA (90% CI)	$\Delta\chi^{2a, b}$	ΔCFI	ΔRMSEA
Single-group CFA						
PL	854.520 (186)	.962	.060 (.056; .064)			
RU	2323.379 (186)	.964	.062 (.059; .064)			
UK	814.666 (186)	.967	.058 (.054; .062)			
US	715.747 (186)	.982	.053 (.049; .057)			
Multi-group CFA						
(1) PL vs. RU vs. UK vs. US						
Configural invariance	4548.319 (744)	.970	.058 (.057; .060)			
Metric invariance	5770.799 (798)	.961	.064 (.063; .066)	|1190.299|	|.009|	|.006|
Threshold invariance	5564.147 (915)	.964	.058 (.057; .060)	|572.725|	|.003|	|.006|
(2) RU vs. UK vs. US						
Configural invariance	3729.319 (558)	.971	.058 (.056; .060)			
Metric invariance	4847.355 (594)	.961	.065 (.064; .067)	|1082.619|	|.010|	|.007|
Threshold invariance	4774.726 (672)	.962	.060 (.059; .062)	|515.841|	|.001|	|.050|
(3) UK vs. US						
Configural invariance	1533.404 (372)	.976	.056 (.053; .059)			
Metric invariance	1578.049 (390)	.976	.055 (.052; .058)	|63.844|	|.000|	|.001|
Threshold invariance	1453.904 (429)	.979	.049 (.046; .052)	|50.103|[c]	|.003|	|.006|

Note. CFA = confirmatory factor analysis; PL = Poland; RU = Russia; UK = United Kingdom; US = United States of America.
[a] With one exception (see below), all χ^2 values are significant, $p \leq .000$. [b] The test was based on robust χ^2 difference testing (calculation based on Asparouhov & Muthen, 2006). [c] χ^2 value is not significant, $p = .11$.

of the latent mean differences of the depression scale indicated significant differences between the Polish sample and U.K., $z = -3.47$, $p = .001$, $d = .16$, and U.S. samples, $z = -3.92$, $p < .001$, $d = .18$, but no difference was found between the Polish and the Russian sample, $z = -0.87$, $p = .387$, $d = .03$. The latent mean of the anxiety scale was significantly higher in the Russian sample than in the Polish sample, $z = -3.47$, $p = .001$, $d = .13$, whereas the differences between the Polish sample and U.K., $z = -1.59$, $p = .113$, $d = .07$, and U.S. samples were not significant, $z = -0.50$, $p = .618$, $d = .02$. The Polish sample scored significantly higher on the latent mean of the stress scale compared with the Russian sample, $z = -4.72$, $p < .001$, $d = .17$, the U.K. sample, $z = -4.49$, $p < .001$, $d = .20$, and the U.S. sample, $z = -3.99$, $p < .001$, $d = .18$.

Table 4
Standardized Factor Loadings of Each Subscale of the DASS-21 Within Each Country

Scale	PL	RU	UK	US
Depression				
dass3	.66	.69	.72	.72
dass5	.59	.68	.68	.74
dass10	.76	.57	.86	.89
dass13	.82	.86	.86	.90
dass16	.83	.76	.85	.86
dass17	.82	.77	.88	.88
dass21	.85	.75	.91	.94
Anxiety				
dass2	.50	.55	.61	.62
dass4	.64	.65	.62	.72
dass7	.68	.73	.64	.72
dass9	.69	.77	.79	.85
dass15	.90	.88	.93	.90
dass19	.72	.69	.69	.78
dass20	.82	.84	.86	.90
Stress				
dass1	.59	.64	.67	.69
dass6	.66	.71	.72	.73
dass8	.74	.82	.81	.88
dass11	.79	.79	.83	.85
dass12	.79	.81	.84	.80
dass14	.71	.76	.64	.74
dass18	.74	.80	.80	.83

Note. PL = Poland; RU = Russia; UK = United Kingdom; US = United States of America.

Russian versus U.K. and U.S. samples. In a second step, the Russian sample's latent mean was fixed to zero. Thus, the latent mean of the Russian sample could be compared with the U.K. and the U.S. sample. On the depression scale, the Russian sample scored significantly higher than the U.K. sample, $z = -3.70$, $p < .001$, $d = .14$, and the U.S. sample, $z = -4.15$, $p < .001$, $d = .15$. The latent mean of the anxiety scale was significantly higher in the Russian sample than in U.K., $z = -6.10$, $p < .001$, $d = .23$, and U.S. samples, $z = -4.26$, $p < .001$, $d = .16$. Finally, the comparisons between stress symptoms in the Russian sample and the U.K. sample, $z = -0.91$, $p = .363$, $d = .03$, and the U.S. sample, $z = -0.44$, $p = .65$, $d = .02$, did not yield significant differences.

U.K. versus U.S. samples. The latent mean of the U.K. sample was fixed to zero to conduct the latent mean comparison between the U.K. and the U.S. sample. U.K. and U.S. samples did not differ from each other significantly neither on the depression, $z = -0.72$, $p = .472$, $d = .03$, and anxiety scales, $z = -0.52$, $p = .604$, $d = .02$, nor on the stress scale, $z = 0.20$, $p = .841$, $d = .01$.

Discussion

The present study examined the measurement invariance of the DASS-21 across randomly selected samples in Poland, Russia, the U.K., and the U.S. Threshold measurement invariance could be established across all countries. Thus statistical requirements for

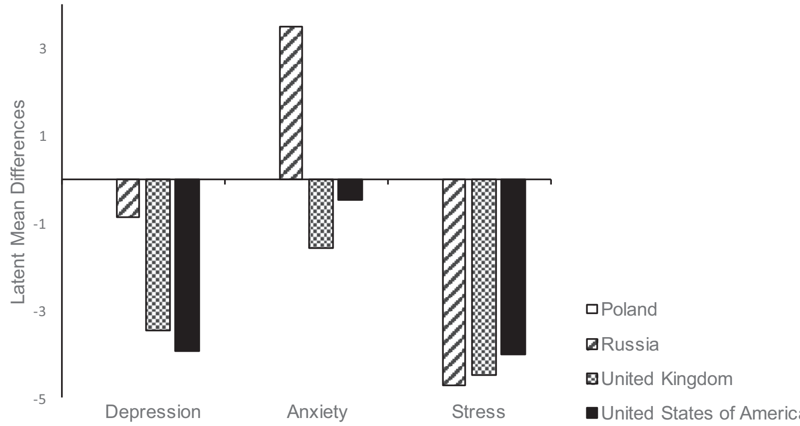

Figure 2. Latent mean comparisons with Poland as the reference country compared with Russia, the U.K. and the U.S.

comparisons of the relationships between latent variables across groups as well as latent mean comparisons are fulfilled.

Scale means were rather low in all samples indicating low symptom levels in the general population. However, the boxplots of the samples demonstrated extreme responses indicating higher symptom levels in a subset of all samples. These subsets might represent the part of the population suffering from more severe mental distress, which would be in line with the high prevalence rates outlined in the introduction. Generally, our findings correspond to the results of Henry and Crawford (2005) and are slightly lower than results from other representative samples (Crawford & Henry, 2003; Lovibond & Lovibond, 1995b; Ronk et al., 2013). This difference could be attributable to the assessment methods. Telephone interviews probably cause more socially desirable behavior and thus result in lower scale scores (Zhang, Kuchinke, & Margraf, 2015).

The results testing the normality assumption were mixed. According to the cut-off scores proposed by Kim (2013) and West et al. (1995), the data were normally distributed, whereas based on Mardia's normalized estimate of multivariate kurtosis the data indicated non-normality in all countries. As the data were left skewed, a moderate non-normality is assumed. Because all analysis were based on WLSMV estimator, such moderate non-normality should not affect the accuracy of the estimation (Flora & Curran, 2004).

Concerning the reliability of the DASS-21, internal consistency was excellent for the total scale and good for the depression, anxiety, and stress scales, respectively. Merely, Cronbach's alpha of the anxiety scale was only acceptable in the Polish sample. These findings, especially the lower internal consistency of the anxiety scale, correspond to the results of previous studies (Antúnez & Vinet, 2012; Bados, Solanas, & Andrés, 2005; Crawford et al., 2011; Daza et al., 2002; Musa et al., 2007; Vasconcelos-Raposo, Fernandes, & Teixeira, 2013).

Within each country, the appropriateness of the theoretically assumed three-factor model was tested. Single-group confirmatory factor analyses showed that the three-factor model fitted the data in all countries. Multigroup confirmatory factor analyses showed that the assumptions of equivalent loadings and thresholds held across all countries. However, it is puzzling that model fit did not decrease from configural to metric and scalar models, as is common in measurement invariance testing using Maximum Likelihood estimation. Instead, χ^2 indicated a better model fit for the more restricted model in all country comparisons. This pattern was also found in the multigroup comparison of the DASS-21 across four racial groups conducted by Norton (2007) who also used WLSMV. Also, ΔCFI and ΔRMSEA did not decrease, nor increase, respectively, as one would expect for the more restricted models. For example, with Poland as the reference country, the CFI was higher in the configural model than in the metric model as expected, but it was also higher in the threshold model compared with the metric model. It is most probable that this pattern is caused by the model identification that is needed in the case of ordered-categorical variables. On each level of measurement invariance testing, different model parameters have to be restricted to identify the model (Millsap & Yun-Tein, 2004; see also the paragraph "Multi-group confirmatory factor analyses to test measurement invariance" in the Method section). Other studies using WLSMV have reported similar patterns (Bieda et al., 2016; Mellor et al., 2015; Norton, 2007), even though the results were not discussed in detail. Further research on multi-Group CFA with ordered-categorical variables should be conducted to better understand the behavior of the Likelihood Ratio Test statistic as well as of the goodness-of-fit indices with ordered-categorical variables. In the present results, all threshold models indicated excellent model fit that did not deviate significantly from the metric models. Hence, these results suggest that relationships between latent variables and latent means can be meaningfully compared across countries.

The results of the latent mean comparisons indicated a clear difference between the two Eastern European and the two Western countries. Whereas the U.S. and the U.K. did not differ in regard to symptoms of depression, anxiety, and stress, both Polish and Russian samples showed more symptoms of depression. However, all effect sizes were very small. High symptom levels of depression are in line with higher prevalence rates of depression found in the Czech Republic, Russia, and Poland (Bobak et al., 2006). One reason for the difference could be the less prosperous economies in Russia and Poland compared with the U.S. and the U.K. Rai, Zitko, Jones, Lynch, and Araya (2013) found that symptoms of depression vary up to 13.5% at the country-level and that symptoms increase with decreasing levels of economic development. More symptoms of anxiety were only reported for the Russian sample compared with the other three countries which did not differ significantly in their anxiety levels. In terms of effect sizes, the Polish and U.S. samples did not differ from the Russian sample, but a small effect was present for the difference between the U.K. and the Russian sample. This finding is in line with very high levels of anxiety found in another Russian sample (Averina et al., 2005) and our findings yield to the fact that this might not be a measurement error. Also prevalence rates of social anxiety have been found to be higher in Russia (Eaton et al., 2008). Such relatively high levels of anxiety might be attributable to rising inequalities in Russia caused by the new globalized neoliberal economic conditions (Burns, 2015). Surprisingly, not the Russian, but the Polish sample reported the highest levels of stress compared with the Russian, U.K., and U.S. sample. A small effect was found compared with the U.K. sample, but effect sizes in comparisons with the Russian and the U.S. samples were very small. Watson (2006) argues that high levels of stress in Poland are linked to a lack of security of employment, lower income, increasing inequality and changes in social relations. It might be particularly burdensome to the Polish population that the country quickly realized its transition to become a democracy and member of the European Union (Carlson, 2016). Now, Poland competes with economically strong countries such as Germany, France, or the U.K., which might increase the pressure and stress levels in people. All of the outlined explanations for latent mean differences across countries should be viewed as a part of a larger picture. Future studies are needed to replicate our findings and test our hypothesis more explicitly to validate or falsify our conclusions.

Generally, such comparisons and hypotheses about cultural differences require prior testing of measurement invariance. The subsequent considerations are very valuable for the comparison of symptoms of mental distress across countries and for the understanding of the social and cultural circumstances behind these differences. This underlines the importance of measurement invariance testing.

Methodological Challenges and General Limitations

Numerous articles and reviews have been published on measurement invariance (e.g., Byrne, 2008; Cheung & Rensvold, 1999; Milfont & Fischer, 2010; Steenkamp & Baumgartner, 1998; van de Schoot, Lugtig, & Hox, 2012; Vandenberg & Lance, 2000). Despite such a large body of research, recommendations for multigroup confirmatory factor analysis with ordered-categorical variables are rare and do not seem to be as straightforward as recommendations for metric variables (Bovaird & Koziol, 2012; Hirschfeld & Von Brachel, 2014; Millsap & Yun-Tein, 2004; B. O. Muthén & Asparouhov, 2002; B. O. Muthén & Christoffersson, 1981). For example, whereas the minimal model identification constraints proposed by Millsap and Yun-Tein (2004) are considered the gold-standard for model identification with ordered-categorical variables, the authors themselves admit that the identification fails in some cases. Bovaird and Koziol (2012) propose a different approach including the restriction of residual variances and Hirschfeld and Von Brachel (2014) do not outline the model identification at all, instead they use the "group.equal" command within the lavaan package (Rosseel, 2012) for the group comparisons. For a clinically oriented researcher such differing approaches might be difficult to evaluate. Furthermore, only recently Sass, Schmitt, and Marsh (2014) gave an overview of the evaluation of model fit with ordered categorical data. Concerning ΔCFI and ΔRMSEA that indicate a meaningful drop of model fit, simulation studies have used maximum-likelihood-estimation (Chen, 2007; Cheung & Rensvold, 2002). It remains unknown whether the cut-off is also a good indicator for models estimated with WLSMV (Hirschfeld & Von Brachel, 2014). Future simulation studies should consider ordered-categorical variables in respect to model identification, estimation, and evaluation in further detail. And most importantly, results should be summarized in comprehensive reviews that can guide future analysis.

Concerning general limitations, only the internal structure of the DASS-21 was analyzed and compared across countries. This leaves other aspects of reliability and validity unanswered and limits the scope of the present study. Nevertheless, as outlined in the introduction, numerous studies have examined the reliability and validity of the DASS-21.

The focus on general population samples across four countries brings about a second general limitation: A part of the population might not be accessible via landline or mobile phones (Brick, Dipko, Presser, Tucker, & Yuan, 2006). Additionally, the response rates indicate that only a small proportion of the generally accessible population agrees to participate in a survey. To deal with accessibility bias and nonresponse, the cases can be weighted according to the respective general population to generalize results (Little & Rubin, 2002). Even though the possibility to weigh cases was available for the present study as well, it was not used in the analyses because weighing cases would mean that the responses of some cases are multiplied and other responses are calculated less than once to fit the gender, age, and regional structure of the population under study. For instance, it is assumed that statistically inflated cases represent the respective population, for example, men aged 34–40 years living in Berlin. However, this assumption could lead to false conclusions based on data with weighted cases (Halbesleben & Whitman, 2013). Nevertheless, only very few studies on the DASS-21 report large samples with the attempt of representativeness (Crawford et al., 2011; Henry & Crawford, 2005; Lovibond & Lovibond, 1995a; Ronk et al., 2013). Thus, despite the limitation of not demonstrating perfect representativeness, from the authors' perspective, we consider the sample large and broad enough to generalize our findings.

A third limitation concerns the method of data collection. The DASS-21 was developed as a paper-pencil questionnaire. In the present study, item responses were assessed verbally in a telephone interview. Altering the survey method might affect social desirability bias (Persoskie, Leyva, & Ferrer, 2014). Another study conducted by our research team (Zhang et al., 2015) compared differences of online, paper-pencil, face-to-face and telephone interviews and the results showed lower severity reports in telephone interviews compared with online or paper-pencil. This would support the hypothesis of the influence of social desirability. However, the effects were smaller than $\eta^2 = 0.04$ and the assessment with the DASS-21 was strongly measurement invariant across all survey methods. Still, it must be kept in mind that responses might be generally lower than they would be in paper-pencil format.

Fourth, measurement invariance is a statistical procedure testing the comparability of an instrument across groups. Yet, it is difficult to disentangle actual cultural differences from the variety of possible biases and artifactual errors in the measurement, such as social norms or data collection methods (Byrne, 2008). Therefore, testing for measurement invariance is only one step toward culturally sensitive measurement instruments. Qualitative research approaches such as cognitive interviewing (Beatty & Willis, 2007) could be used in the translation process to address possible construct and item bias, but also in post hoc analysis to explore reasons for noninvariance.

Conclusion

Facing the burden of an increasing number of mental disorders, researchers worldwide need to work together to develop, exchange, and evaluate prevention and treatment strategies. For collaboration in research and practice, reliable and valid assessment instruments that enable the measurement and comparison of psychological symptoms are fundamental. Cross-cultural research and practice should address additional methodological questions concerning measurement invariance of the assessment instruments as well as questions regarding real cultural differences. This study presents tests of measurement invariance of the DASS-21 that was shown to be threshold measurement invariant in Poland, Russia, the U.K., and the U.S. The analyses of relationships between latent variables as well as latent mean comparisons are allowed based on the full scales across all investigated countries. Future studies in cross-cultural settings need to test measurement invariance in other clinical assessment instruments to overcome the challenge of assessing psychological symptoms across nations and to avoid bias due to the interpretation of distorted observed scale means. More methodological research on measurement invariance testing with ordered-categorical variables is needed to clarify and standardize the procedures of the analysis.

References

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment, 10,* 176–181. http://dx.doi.org/10.1037/1040-3590.10.2.176

Antúnez, Z., & Vinet, E. V. (2012). Escalas de depresión, ansiedad y estrés (DASS–21): Validación de la versión abreviada en estudiantes universitarios chilenos [Depression Anxiety Stress Scales (DASS–21): Validation of the abbreviated version in Chilean university students]. *Terapia Psicológica, 30,* 49–55. http://dx.doi.org/10.4067/S0718-48082012000300005

Apóstolo, J. L. A., Mendes, A. C., & Azeredo, Z. A. (2006). Adaptación para la lengua portuguesa de la depression, anxiety, and stress scale (DASS) [Adaptation to Portuguese of the Depression, Anxiety, and Stress Scales]. *Revista Latino-Americana de Enfermagem, 14,* 863–871. http://dx.doi.org/10.1590/S0104-11692006000600006

Asghari, A., Saed, F., & Dibajnia, P. (2008). Psychometric properties of the Depression Anxiety Stress Scales-21 (DASS-21) in a non-clinical Iranian sample. *International Journal of Psychology, 2,* 82–102.

Asparouhov, T., & Muthén, B. O. (2006). *Robust chi square difference testing with mean and variance adjusted test statistics: Mplus Web Notes: No. 10*. Retrieved from http://www.statmodel.com/download/webnotes/webnote10.pdf

Asparouhov, T., & Muthén, B. O. (2010). *Weighted least squares estimation with missing data*. Retrieved March 12, 2016, from http://www.statmodel.com/download/GstrucMissingRevision.pdf

Averina, M., Nilssen, O., Brenn, T., Brox, J., Arkhipovsky, V. L., & Kalinin, A. G. (2005). Social and lifestyle determinants of depression, anxiety, sleeping disorders and self-evaluated quality of life in Russia—A population-based study in Arkhangelsk. *Social Psychiatry and Psychiatric Epidemiology, 40,* 511–518. http://dx.doi.org/10.1007/s00127-005-0918-x

Bados, A., Solanas, A., & Andrés, R. (2005). Psychometric properties of the Spanish version of Depression, Anxiety, and Stress Scales (DASS). *Psicothema, 17,* 679–683.

Baxter, A. J., Vos, T., Scott, K. M., Norman, R. E., Flaxman, A. D., Blore, J., & Whiteford, H. A. (2014). The regional distribution of anxiety disorders: Implications for the Global Burden of Disease Study, 2010. *International Journal of Methods in Psychiatric Research, 23,* 422–438. http://dx.doi.org/10.1002/mpr.1444

Bayani, A. A. (2010). Reliability and preliminary evidence of validity of a Farsi version of the Depression Anxiety Stress Scales. *Perceptual and Motor Skills, 111,* 107–114. http://dx.doi.org/10.2466/08.13.PMS.111.4.107-114

Bayram, N., & Bilgel, N. (2008). The prevalence and socio-demographic correlations of depression, anxiety and stress among a group of university students. *Social Psychiatry and Psychiatric Epidemiology, 43,* 667–672. http://dx.doi.org/10.1007/s00127-008-0345-x

Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly, 71,* 287–311. http://dx.doi.org/10.1093/poq/nfm006

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, 13,* 186–203. http://dx.doi.org/10.1207/s15328007sem1302_2

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246. http://dx.doi.org/10.1037/0033-2909.107.2.238

Berry, J. W. (1969). On cross-cultural comparability. *International Journal of Psychology, 4,* 119–128. http://dx.doi.org/10.1080/00207596908247261

Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross-cultural psychology: Research and applications* (2nd ed.). New York, NY: Cambridge University Press.

Bieda, A., Hirschfeld, G., Schönfeld, P., Brailovskaia, J., Zhang, X. C., & Margraf, J. (2016). Universal happiness? Cross-cultural measurement invariance of scales assessing positive mental health. *Psychological Assessment*. Advance online publication. http://dx.doi.org/10.1037/pas0000353

Bobak, M., Pikhart, H., Pajak, A., Kubinova, R., Malyutina, S., Sebakova, H., . . . Marmot, M. (2006). Depressive symptoms in urban population samples in Russia, Poland and the Czech Republic. *The British Journal of Psychiatry, 188,* 359–365. http://dx.doi.org/10.1192/bjp.188.4.359

Borsboom, D. (2006). When does measurement invariance matter? *Medical Care, 44*(11, Suppl. 3), S176–S181. http://dx.doi.org/10.1097/01.mlr.0000245143.08679.cc

Bottesi, G., Ghisi, M., Altoè, G., Conforti, E., Melli, G., & Sica, C. (2015). The Italian version of the Depression Anxiety Stress Scales-21: Factor structure and psychometric properties on community and clinical samples. *Comprehensive Psychiatry, 60,* 170–181. http://dx.doi.org/10.1016/j.comppsych.2015.04.005

Bovaird, A. J., & Koziol, N. A. (2012). Measurement models for ordered-categorical indicators. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 495–511). New York, NY: Guilford Press.

Bowden, A., & Fox-Rushby, J. A. (2003). A systematic and critical review of the process of translation and adaptation of generic health-related quality of life measures in Africa, Asia, Eastern Europe, the Middle East, South America. *Social Science & Medicine, 57,* 1289–1306. http://dx.doi.org/10.1016/S0277-9536(02)00503-8

Brick, J. M., Dipko, S., Presser, S., Tucker, C., & Yuan, Y. (2006). Nonresponse bias in a Dual Frame sample of cell and landline numbers. *Public Opinion Quarterly, 70,* 780–793. http://dx.doi.org/10.1093/poq/nfl031

Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behaviour Research and Therapy, 35,* 79–89. http://dx.doi.org/10.1016/S0005-7967(96)00068-X

Burns, J. K. (2015). Poverty, inequality and a political economy of mental health. *Epidemiology and Psychiatric Sciences, 24,* 107–113. http://dx.doi.org/10.1017/S2045796015000086

Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema, 20,* 872–882.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York, NY: Routledge.

Byrne, B. M. (2016). Adaptation of assessment scales in cross-national research: Issues, guidelines, and caveats. *International Perspectives in Psychology: Research, Practice, Consultation, 5,* 51–65. http://dx.doi.org/10.1037/ipp0000042

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

Byrne, B. M., Stewart, S. M., Kennard, B. D., & Lee, P. W. H. (2007). The Beck Depression Inventory-II: Testing for measurement equivalence and factor mean differences across Hong Kong and American adolescents. *International Journal of Testing, 7,* 293–309. http://dx.doi.org/10.1080/15305050701438058

Canel-Çınarbaş, D., Cui, Y., & Lauridsen, E. (2011). Cross-cultural validation of the Beck Depression Inventory-II across U.S. and Turkish samples. *Measurement and Evaluation in Counseling and Development, 44,* 77–91. http://dx.doi.org/10.1177/0748175611400289

Carlson, P. (2016). Trust and health in Eastern Europe: Conceptions of a new society. *International Journal of Social Welfare, 25,* 69–77. http://dx.doi.org/10.1111/ijsw.12165

Chan, R. C. K., Xu, T., Huang, J., Wang, Y., Zhao, Q., Shum, D. H. K., . . . Potangaroa, R. (2012). Extending the utility of the Depression Anxiety Stress Scale by examining its psychometric properties in Chinese settings. *Psychiatry Research, 200,* 879–883. http://dx.doi.org/10.1016/j.psychres.2012.06.041

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14,* 464–504. http://dx.doi.org/10.1080/10705510701301834

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology, 95,* 1005–1018. http://dx.doi.org/10.1037/a0013193

Cheung, F. M. (2012). Mainstreaming culture in psychology. *American Psychologist, 67,* 721–730. http://dx.doi.org/10.1037/a0029876

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25,* 1–27. http://dx.doi.org/10.1177/014920639902500101

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233–255. http://dx.doi.org/10.1207/S15328007SEM0902_5

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100,* 316–336. http://dx.doi.org/10.1037/0021-843X.100.3.316

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd editio). Hillsdale, NJ: Erlbaum.

Corporation, I. B. M. (2011). SPSS Statistics (Version 20.0) [Computer software]. Amonk, NY: IBM SPSS Statistics.

Crawford, J., Cayley, C., Lovibond, P. F., Wilson, P. H., & Hartley, C. (2011). Percentile norms and accompanying interval estimates from an Australian general adult population sample for self-report mood scales (BAI, BDI, CRSD, CES-D, DASS, DASS-21, STAI-X, STAI-Y, SRDS, and SRAS). *Australian Psychologist, 46,* 3–14. http://dx.doi.org/10.1111/j.1742-9544.2010.00003.x

Crawford, J. R., & Henry, J. D. (2003). The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology, 42,* 111–131. http://dx.doi.org/10.1348/014466503321903544

Crockett, L. J., Randall, B. A., Shen, Y.-L., Russell, S. T., & Driscoll, A. K. (2005). Measurement equivalence of the center for epidemiological studies depression scale for Latino and Anglo adolescents: A national study. *Journal of Consulting and Clinical Psychology, 73,* 47–58. http://dx.doi.org/10.1037/0022-006X.73.1.47

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297–334. http://dx.doi.org/10.1007/BF02310555

Dahm, J., Wong, D., & Ponsford, J. (2013). Validity of the Depression Anxiety Stress Scales in assessing depression and anxiety following traumatic brain injury. *Journal of Affective Disorders, 151,* 392–396. http://dx.doi.org/10.1016/j.jad.2013.06.011

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology, 40,* 55–75. http://dx.doi.org/10.1146/annurev-soc-071913-043137

Davies, G., Caputi, P., Skarvelis, M., & Ronan, N. (2015). The Depression, Anxiety, and Stress Scales: Reference data from a large psychiatric outpatient sample. *Australian Journal of Psychology, 67,* 97–104. http://dx.doi.org/10.1111/ajpy.12069

Daza, P., Novy, D. M., Stanley, M. A., & Averill, P. (2002). The Depression Anxiety Stress Scale-21: Spanish translation and validation with a Hispanic sample. *Journal of Psychopathology and Behavioral Assessment, 24,* 195–205. http://dx.doi.org/10.1023/A:1016014818163

Dere, J., Watters, C. A., Yu, S. C-M., Bagby, R. M., Ryder, A. G., Harkness, K. L., & Harkness, K. L. (2015). Cross-cultural examination of measurement invariance of the Beck Depression Inventory-II. *Psychological Assessment, 27,* 68–81. http://dx.doi.org/10.1037/pas0000026

Eaton, W. W., Martins, S. S., Nestadt, G., Bienvenu, O. J., Clarke, D., & Alexandre, P. (2008). The burden of mental disorders. *Epidemiologic Reviews, 30,* 1–14. http://dx.doi.org/10.1093/epirev/mxn011

European Social Survey. (2012). *Appendix A1 Education, ESS6-2012 Ed. 2.0*. Retrieved March 12, 2016, from http://www.europeansocialsurvey.org/docs/round6/survey/ESS6_appendix_a1_e02_0.pdf

Eurostat. (2007). *Task force on core social variables*. Retrieved December 3, 2015, from http://ec.europa.eu/eurostat/documents/3859598/5901513/KS-RA-07-006-EN.PDF/71481ffb-771a-489b-a749-1a055c0247d4

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9,* 466–491. http://dx.doi.org/10.1037/1082-989X.9.4.466

Gloster, A. T., Rhoades, H. M., Novy, D., Klotsche, J., Senior, A., Kunik, M., . . . Stanley, M. A. (2008). Psychometric properties of the Depression Anxiety and Stress Scale-21 in older primary care patients. *Journal of Affective Disorders, 110,* 248–259. http://dx.doi.org/10.1016/j.jad.2008.01.023

Gomez, R., Summers, M., Summers, A., Wolf, A., & Summers, J. (2014). Depression Anxiety Stress Scales-21: Measurement and structural invariance across ratings of men and women. *Assessment, 21,* 418–426. http://dx.doi.org/10.1177/1073191113514106

Gong, X., Xie, X., Xu, R., & Luo, Y. (2010). Psychometric properties of the Chinese versions of DASS-21 in Chinese college students. *Chinese Journal of Clinical Psychology, 18,* 443–446.

Halbesleben, J. R. B., & Whitman, M. V. (2013). Evaluating survey quality in health services research: A decision framework for assessing nonresponse bias. *Health Services Research, 48,* 913–930. http://dx.doi.org/10.1111/1475-6773.12002

Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3–38). Mahwah, NJ: Erlbaum.

He, J., & van de Vijver, F. J. R. (2012). Bias and equivalence in cross-cultural research bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture, 2*. http://dx.doi.org/10.9707/2307-0919.1111

Hekimoglu, L., Altun, Z. O., Kaya, E. Z., Bayram, N., & Bilgel, N. (2012). Psychometric properties of the Turkish version of the 42 item Depression Anxiety Stress Scale (DASS-42) in a clinical sample. *International Journal of Psychiatry in Medicine, 44,* 183–198. http://dx.doi.org/10.2190/PM.44.3.a

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 44,* 227–239. http://dx.doi.org/10.1348/014466505X29657

Hirschfeld, G., & Von Brachel, R. (2014). Multiple-group confirmatory factor analysis in R—A tutorial in measurement invariance with continuous and ordinal. *Practical Assessment, Research & Evaluation, 19,* 1–11.

Hoe, M., & Brekke, J. (2009). Testing the cross-ethnic construct validity of the Brief Symptom Inventory. *Research on Social Work Practice, 19,* 93–103. http://dx.doi.org/10.1177/1049731508317285

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. http://dx.doi.org/10.1080/10705519909540118

International Test Commission. (2005). *International guidelines on test adaptation*. https://www.intestcom.org/

International Test Commission. (2016). *International Test Commission*. Retrieved September 8, 2016, from https://www.intestcom.org/

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36,* 409–426. http://dx.doi.org/10.1007/BF02291366

Jovanovic, V., Gavrilov-Jerkovic, V., Zuljevic, D., & Brdaric, D. (2014). Psychometric evaluation of the depression anxiety stress scales-21 (DASS-21) in a Serbian student sample. *Psihologija (Beograd), 47,* 93–112. http://dx.doi.org/10.2298/PSI1401093J

Kessler, R. C., Aguilar-Gaxiola, S., Alonso, J., Chatterji, S., Lee, S., Ormel, J., . . . Wang, P. S. (2009). The global burden of mental disorders: An update from the WHO World Mental Health (WMH) surveys. *Epidemiologia e Psichiatria Sociale, 18,* 23–33. http://dx.doi.org/10.1017/S1121189X00001421

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month *DSM–IV* disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62,* 617–627. http://dx.doi.org/10.1001/archpsyc.62.6.617

Kim, H. Y. (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics, 38,* 52–54. http://dx.doi.org/10.5395/rde.2013.38.1.52

Kirmayer, L. J. (2001). Cultural variations in the clinical presentation of depression and anxiety: Implications for diagnosis and treatment. *The Journal of Clinical Psychiatry, 62*(Suppl. 13), 22–28–30.

Kok, T., de Haan, H. A., van der Meer, M., Najavits, L. M., & De Jong, C. A. J. (2015). Screening of current post-traumatic stress disorder in patients with substance use disorder using the Depression, Anxiety and Stress Scale (DASS-21): A reliable and convenient measure. *European Addiction Research, 21,* 71–77. http://dx.doi.org/10.1159/000365283

Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data*. New York, NY: Wiley. http://dx.doi.org/10.1002/9781119013563

López, S. R., & Guarnaccia, P. J. (2000). Cultural psychopathology: Uncovering the social world of mental illness. *Annual Review of Psychology, 51,* 571–598. http://dx.doi.org/10.1146/annurev.psych.51.1.571

Lovibond, P. F. (2015). *DASS translations*. Retrieved November 19, 2015, from http://www2.psy.unsw.edu.au/groups/dass/translations.htm

Lovibond, P. F., & Lovibond, S. H. (1995a). *Manual for the Depression Anxiety Stress Scales* (2nd ed.). Book, Sydney: Psychology Foundation.

Lovibond, P. F., & Lovibond, S. H. (1995b). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy, 33,* 335–343. http://dx.doi.org/10.1016/0005-7967(94)00075-U

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57,* 519–530. http://dx.doi.org/10.1093/biomet/57.3.519

Mellor, D., Vinet, E. V., Xu, X., Mamat, N. H. B., Richardson, B., & Román, F. (2015). Factorial invariance of the DASS-21 among adolescents in four countries. *European Journal of Psychological Assessment, 31,* 138–142. http://dx.doi.org/10.1027/1015-5759/a000218

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research, 3,* 111–121.

Millsap, R. E., & Kwok, O.-M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods, 9,* 93–115. http://dx.doi.org/10.1037/1082-989X.9.1.93

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

Murray, C. J. L., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., . . . Lopez, A. D. (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet, 380,* 2197–2223. http://dx.doi.org/10.1016/S0140-6736(12)61689-4

Musa, R., Fadzil, M. A., & Zain, Z. (2007). Translation, validation, and psychometric properties of Bahasa Malaysia version of the Depression, Anxiety, and Stress Scales (DASS). *ASEAN Journal of Psychiatry, 8,* 82–89.

Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49,* 115–132. http://dx.doi.org/10.1007/BF02294210

Muthén, B. O., & Asparouhov, T. (2002). *Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus*. Retrieved March 11, 2016, from https://www.statmodel.com/download/webnotes/CatMGLong.pdf

Muthén, B. O., & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika, 46,* 407–419. http://dx.doi.org/10.1007/BF02293798

Muthén, L. K., & Muthén, B. O. (2012). *Mplus: Statistical analysis with latent variables; User's guide* (7th ed.). Los Angeles, CA: Author.

Muthén, L. K., & Muthén, B. O. (2015). *Mplus*. Los Angeles, CA: Author.

Ng, F., Trauer, T., Dodd, S., Callaly, T., Campbell, S., & Berk, M. (2007). The validity of the 21-item version of the Depression Anxiety Stress Scales as a routine clinical outcome measure. *Acta Neuropsychiatrica, 19,* 304–310. http://dx.doi.org/10.1111/j.1601-5215.2007.00217.x

Nieuwenhuijsen, K., de Boer, A. G. E. M., Verbeek, J. H. A. M., Blonk, R. W. B., & van Dijk, F. J. H. (2003). The Depression Anxiety Stress Scales (DASS): Detecting anxiety disorder and depression in employees absent from work because of mental health problems. *Occupational and*

Environmental Medicine, 60, i77–i82. http://dx.doi.org/10.1136/oem.60.suppl_1.i77

Nilges, P., & Essau, C. (2015). Depression, Anxiety, and Stress Scales: The DASS - A screening not only for patients with chronic pain [Die Depressions-Angst-Stress-Skalen: Der DASS—Ein Screeningverfahren nicht nur für Schmerzpatienten]. *Schmerz, 29,* 649–657. http://dx.doi.org/10.1007/s00482-015-0019-z

Norton, P. J. (2007). Depression Anxiety and Stress Scales (DASS-21): Psychometric analysis across four racial groups. *Anxiety, Stress, & Coping, 20,* 253–265. http://dx.doi.org/10.1080/10615800701309279

Nuevo, R., Dunn, G., Dowrick, C., Vázquez-Barquero, J. L., Casey, P., Dalgard, O. S., . . . Ayuso-Mateos, J. L. (2009). Cross-cultural equivalence of the Beck Depression Inventory: A five-country analysis from the ODIN study. *Journal of Affective Disorders, 114*(1–3), 156–162. http://dx.doi.org/10.1016/j.jad.2008.06.021

Oei, T. P. S., Sawang, S., Goh, Y. W., & Mukhtar, F. (2013). Using the Depression Anxiety Stress Scale 21 (DASS-21) across cultures. *International Journal of Psychology, 48,* 1018–1029. http://dx.doi.org/10.1080/00207594.2012.755535

Page, A. C., Hooke, G. R., & Morrison, D. L. (2007). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in depressed clinical samples. *British Journal of Clinical Psychology, 46,* 283–297. http://dx.doi.org/10.1348/014466506X158996

Pendergast, L. L., Youngstrom, E. A., Brown, C., Jensen, D., Abramson, L. Y., & Alloy, L. B. (2015). Structural invariance of General Behavior Inventory (GBI) scores in Black and White young adults. *Psychological Assessment, 27,* 21–30. http://dx.doi.org/10.1037/pas0000020

Persoskie, A., Leyva, B., & Ferrer, R. A. (2014). Mode effects in assessing cancer worry and risk perceptions: Is social desirability bias at play? *Medical Decision Making, 34,* 583–589. http://dx.doi.org/10.1177/0272989X14527173

Poortinga, Y. H. (1995). Cultural bias in assessment: Historical and thematic issues. *European Journal of Psychological Assessment, 11,* 140–146. http://dx.doi.org/10.1027/1015-5759.11.3.140

Rai, D., Zitko, P., Jones, K., Lynch, J., & Araya, R. (2013). Country- and individual-level socioeconomic determinants of depression: Multilevel cross-national comparison. *The British Journal of Psychiatry, 202,* 195–203. http://dx.doi.org/10.1192/bjp.bp.112.112482

Ramli, M., Ramli, R., Kartini, A., & Rosnani, S. (2011). Concurrent validity of the depression and anxiety components in the Bahasa Malaysia version of the Depression, Anxiety, and Stress Scales (DASS). *ASEAN Journal of Psychiatry, 12*.

Ramli, M., Salmiah, M. A., & Nurul Ain, M. (2009). Validation and psychometric properties of Bahasa Malaysia version of the Depression, Anxiety, and Stress Scales (DASS) among diabetic patients. *MJP Online Early,* 1–6.

R Development Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org

Revelle, W. (2015). *Psych: Procedures for personality and psychological research*. Retrieved from http://personality-project.org/r

Ronk, F. R., Korman, J. R., Hooke, G. R., & Page, A. C. (2013). Assessing clinical significance of treatment outcomes using the DASS-21. *Psychological Assessment, 25,* 1103–1110. http://dx.doi.org/10.1037/a0033100

Rosseel, Y. (2012). Lavaan: An R package for structural equation. *Journal of Statistical Software, 48,* 1–36. http://dx.doi.org/10.18637/jss.v048.i02

Ryan, T. A., Bailey, A., Fearon, P., & King, J. (2013). Factorial invariance of the Patient Health Questionnaire and Generalized Anxiety Disorder Questionnaire. *British Journal of Clinical Psychology, 52,* 438–449. http://dx.doi.org/10.1111/bjc.12028

Ryder, A. G., Yang, J., Zhu, X., Yao, S., Yi, J., Heine, S. J., & Bagby, R. M. (2008). The cultural shaping of depression: Somatic symptoms in China, psychological symptoms in North America? *Journal of Abnormal Psychology, 117,* 300–313. http://dx.doi.org/10.1037/0021-843X.117.2.300

Sahebi, A., Asghari, M. J., & Salari, R. S. (2005). Validation of Depression, Anxiety, and Stress Scale (DASS-21) for an Iranian population. *Journal of Iranian Psychologists, 1,* 299–310.

Sardá, J., Jr., Nicholas, M. K., Pimenta, C. A., & Asghari, A. (2008). Psychometric properties of the DASS-Depression Scale among a Brazilian population with chronic pain. *Journal of Psychosomatic Research, 64,* 25–31. http://dx.doi.org/10.1016/j.jpsychores.2007.05.015

Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment, 29,* 347–363. http://dx.doi.org/10.1177/0734282911406661

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling, 21,* 167–180. http://dx.doi.org/10.1080/10705511.2014.882658

Satorra, A., & Bentler, P. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66,* 507–514. http://dx.doi.org/10.1007/BF02296192

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment, 29,* 304–321. http://dx.doi.org/10.1177/0734282911406653

Severino, G. A., & Haynes, W. D. G. (2010). Development of an Italian version of the Depression Anxiety Stress Scales. *Psychology, Health, & Medicine, 15,* 607–621. http://dx.doi.org/10.1080/13548506.2010.498889

Sinclair, S. J., Siefert, C. J., Slavin-Mulford, J. M., Stein, M. B., Renna, M., & Blais, M. A. (2012). Psychometric evaluation and normative data for the depression, anxiety, and stress scales-21 (DASS-21) in a nonclinical sample of U.S. adults. *Evaluation & the Health Professions, 35,* 259–279. http://dx.doi.org/10.1177/0163278711424282

Singh, K., Junnarkar, M., & Sharma, S. (2015). Anxiety, stress, depression, and psychosocial functioning of Indian adolescents. *Indian Journal of Psychiatry, 57,* 367–374. http://dx.doi.org/10.4103/0019-5545.171841

Smith, S., Sullivan, K., Hopkins, W., & Douglas, J. (2004). Frequency of insomnia report in patients with obstructive sleep apnoea hypopnea syndrome (OSAHS). *Sleep Medicine, 5,* 449–456. http://dx.doi.org/10.1016/j.sleep.2004.03.005

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25,* 78–90. http://dx.doi.org/10.1086/209528

Steiger, J. H., & Lind, J. C. (1980). Statistically based tests for the number of common factors. In Annual meeting of the Psychometric Society, Iowa City, IA.

Steinmetz, H. (2013). Analyzing observed composite differences across groups. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 9,* 1–12. http://dx.doi.org/10.1027/1614-2241/a000049

Szabó, M. (2011). The emotional experience associated with worrying: Anxiety, depression, or stress? *Anxiety, Stress, & Coping, 24,* 91–105. http://dx.doi.org/10.1080/10615801003653430

Taylor, R., Lovibond, P. F., Nicholas, M. K., Cayley, C., & Wilson, P. H. (2005). The utility of somatic items in the assessment of depression in patients with chronic pain: A comparison of the Zung Self-Rating Depression Scale and the Depression Anxiety Stress Scales in chronic pain and clinical and community samples. *The Clinical Journal of Pain, 21,* 91–100. http://dx.doi.org/10.1097/00002508-200501000-00011

The American Association for Public Opinion Research. (2016). *Standard definitions - Final dispositions of case codes and outcome rates for surveys* (9th ed.). http://doi.org/10.1111/J.1600-6143.2005.00812.X

Tonsing, K. N. (2014). Psychometric properties and validation of Nepali version of the Depression Anxiety Stress Scales (DASS-21). *Asian*

Journal of Psychiatry, 8, 63–66. http://doi.org/10.1016/j.ajp.2013.11.001

Torres, L., Miller, M. J., & Moore, K. M. (2013). Factorial invariance of the Brief Symptom Inventory-18 (BSI-18) for adults of Mexican descent across nativity status, language format, and gender. *Psychological Assessment, 25,* 300–305. http://dx.doi.org/10.1037/a0030436

Tran, T. D., Tran, T., & Fisher, J. (2013). Validation of the Depression Anxiety Stress Scales (DASS) 21 as a screening instrument for depression and anxiety in a rural community-based cohort of northern Vietnamese women. *BMC Psychiatry, 13,* 24. http://dx.doi.org/10.1186/1471-244X-13-24

Uncu, Y., Bayram, N., & Bilgel, N. (2007). Job related affective well-being among primary health care physicians. *European Journal of Public Health, 17,* 514–519. http://dx.doi.org/10.1093/eurpub/ckl264

UNESCO Institute for Statistics. (2013). The International Standard Classification of Education 2011. *Comparative Social Research*. http://dx.doi.org/10.15220/978-92-9189-123-8-en

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,* 4–70. http://dx.doi.org/10.1177/109442810031002

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9,* 486–492. http://dx.doi.org/10.1080/17405629.2012.686740

van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research* (Vol. 1). Thousand Oaks, CA: SAGE Publication.

van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology/Revue Européenne de Psychologie Appliquée, 54,* 119–135. http://dx.doi.org/10.1016/j.erap.2003.12.004

Vasconcelos-Raposo, J., Fernandes, H. M., & Teixeira, C. M. (2013). Factor structure and reliability of the Depression, Anxiety and Stress Scales in a large Portuguese community sample. *The Spanish Journal of Psychology, 16,* E10. http://dx.doi.org/10.1017/sjp.2013.15

Vignola, R. C. B., & Tucci, A. M. (2014). Adaptation and validation of the Depression, Anxiety and Stress Scale (DASS) to Brazilian Portuguese. *Journal of Affective Disorders, 155,* 104–109. http://dx.doi.org/10.1016/j.jad.2013.10.031

Wang, K., Shi, H. S., Geng, F. L., Zou, L. Q., Tan, S. P., Wang, Y., . . . Chan, R. C. K. (2016). Cross-cultural validation of the Depression Anxiety Stress Scale-21 in China. *Psychological Assessment, 28,* e88–e100. http://dx.doi.org/10.1037/pas0000207

Watson, P. (2006). Stress and social change in Poland. *Health & Place, 12,* 372–382. http://dx.doi.org/10.1016/j.healthplace.2005.02.007

Weiss, M. G., & Somma, D. (2007). Explanatory models in psychiatry. In D. Bhugra & K. Bhui (Eds.), *Textbook of Cultural Psychiatry* (pp. 127–140). Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511543609.012

West, S. G., Finch, J. F., & Curran, P. (1995). Structural equation models with non-normal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). Thousand Oaks, CA: Sage.

Whisman, M. a, Judd, C. M., Whiteford, N. T., & Gelhorn, H. L. (2013). Measurement invariance of the Beck Depression Inventory-2nd ed. (BDI-II) across gender, race, and ethnicity in college students. *Assessment, 20,* 419–28. http://doi.org/10.1177/1073191112460273

Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., . . . the ISPOR Task Force for Translation and Cultural Adaptation. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in Health, 8,* 94–104. http://dx.doi.org/10.1111/j.1524-4733.2005.04054.x

Wilson, J. K., & Rapee, R. M. (2005). The interpretation of negative social events in social phobia with versus without comorbid mood disorder. *Journal of Anxiety Disorders, 19,* 245–274. http://dx.doi.org/10.1016/j.janxdis.2004.03.003

Wittchen, H.-U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., . . . Steinhausen, H.-C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology, 21,* 655–679. http://dx.doi.org/10.1016/j.euroneuro.2011.07.018

Zhang, X. C., Kuchinke, L., & Margraf, J. (2015). *Erhebungsmethode: Beeinflussen Face-to-face, Online oder Telefon die Outcome-Variablen?* [Assessment method: Do face-to-face, online or telephone influence the outcome variable?]. In Presentation on the 9th Workshopcongress for Clinical Psychology and Psychotherapy of the DGPs Section for Clinical Psychology and Psychotherapy, Dresden, Germany.

Received December 4, 2015
Revision received December 2, 2016
Accepted December 7, 2016 ■