



Testing Measurement Invariance of the Positive and Negative Affect Schedule (PANAS) in American and Arab University Students

Rachel C. Davis¹ · M. Alejandra Arce¹ · Kaitlyn E. Tobin¹ · Isabella M. Palumbo¹ · Michael Chmielewski² · Ahmed M. Megreya³ · Robert D. Latzman¹ 

Accepted: 7 October 2020/Published online: 04 November 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Affect describes any feelings, emotions, or moods that a person experiences and is generally divided into two broad dimensions—positive affect and negative affect. The most widely used measure of affect, the Positive and Negative Affect Schedule (PANAS), has recently been translated to Arabic (see Appendix), yet the psychometric equivalence of this adapted measure is not fully understood. Using a series of multigroup confirmatory factor analyses, the current study examined measurement invariance of the English and adapted Arabic versions of the PANAS among 979 American and 1470 Arab university students. Although the two-factor structure of the 20-item PANAS was observed in both groups (configural variance), results did not support full invariance of factor loadings (metric invariance). A partial metric invariance model, however, revealed invariant loadings for all positive affect items and all but four negative affect items; dissimilar factor loadings emerged between groups for *irritable*, *nervous*, *scared*, and *jittery*. Evidence did not support scalar invariance of the 16 metric-invariant items, with only ten items demonstrating equivalent intercepts across groups. Finally, tests for strict invariance indicated nine of the ten scalar-invariant items had equivalent residual variances across groups.

All told, results suggest that the PANAS is partially invariant in American and Arab groups.

Keywords Positive and Negative Affect Schedule · Measurement invariance · Cross-cultural · Affect

✉ Robert D. Latzman
rlatzman@gsu.edu

¹ Department of Psychology, Georgia State University, PO Box 5010, Atlanta, GA 30302-5010, USA

² Southern Methodist University, Dallas, TX, USA

³ Qatar University, Doha, Qatar

The term *affect* describes any feelings, emotions, or moods that a person experiences and is generally divided into two broad, largely orthogonal dimensions—positive affect and negative affect (Watson et al. 1988a, b). Positive affect describes a bipolar dimension on which high positive affect is characterized by high energy, enthusiasm, and pleasurable engagement and low positive affect by feelings such as lethargy. In contrast, negative affect describes the extent to which an individual experiences unpleasurable emotions, with high negative affect characterized by feelings such as anger, distress, and fear and low negative affect characterized by a state of calmness. Affect dimensions have been linked to a range of psychobiological (Dockray and Steptoe 2010) and social (Bryan et al. 1998; Bodenhausen et al. 1994) processes and outcomes and are considered reliable predictors of psychopathology, particularly anxiety and depression (Watson et al. 1988a). Thus, measuring affect is important for advancing our understanding of emotion and its many outcomes.

The Positive and Negative Affect Schedule (PANAS; Watson et al. 1988b) is the most widely used measure of affect, perhaps due to its brevity, reliability, and ability to distinguish between anxiety and depression (Watson et al. 1988a). Since its development and initial validation with samples of university students, university-based employees, adult community members, and psychiatric patients (Watson et al. 1988b), the PANAS has been used to measure affect in both community (e.g., Merz et al. 2013; Crawford and Henry 2004) and clinical (e.g., Díaz-García et al. 2020; Estévez-López et al. 2016) samples and has demonstrated adequate psychometric properties across samples. Such characteristics make this measure useful and relevant for assessing mood, within both normal and abnormal ranges, as well as understanding variation in psychopathology. The PANAS was developed to assess the well-established, largely independent, two dimensions of affect, positive and negative affect, each operationalized by ten items with minimal cross-loading (Watson et al. 1988a, b). The psychometric properties of this measure have been widely examined and the proposed two-factor structure of affect is well-established in the literature (e.g., Crawford and Henry 2004; Serafini et al. 2016; Tuccitto et al. 2010).

Although the PANAS was originally developed in English, it has been translated into a variety of languages, including Spanish (Joiner et al. 1997), Italian (Terraciano et al. 2003), Hindi (Pandey and Srivastava 2008), and many others. Recently, Megreya et al. (2016) reported on an Arabic translation (see Appendix). It is not yet clear, though, whether the measurement of affect in this adapted Arabic version is parallel to its measurement in the original English version. That is, research has yet to investigate whether PANAS items contribute equally to positive and negative affect across English-speaking and Arabic-speaking groups, whether mean levels of each item are the same across groups, and whether residual variances of each item are the same across groups. Although the majority of psychometric research involving the cross-cultural adaptation of measures has historically focused on reliability and validity, it is also important to consider measurement invariance. Indeed, as described in greater detail below, measurement invariance provides statistical confirmation that the measure has the same meaning across groups (Putnick and Bornstein 2016). Without evidence of measurement invariance, it is difficult to determine whether differences between groups are a result of true differences in the constructs measured rather than simply reflecting measurement error as a result of measurement noninvariance across groups.

Levels of Measurement Invariance

Although there are multiple methods and levels of invariance one can examine—each with unique implications—measurement invariance is broadly concerned with the generalization of measurement properties among two or more groups. Establishing measurement invariance is thus an important first step when seeking to examine between-group differences. If measurement invariance is not established but simply assumed in such examinations, then the conclusions that can be drawn are limited and likely biased. That is, the observed group differences may not indicate true group differences, but rather suggest that the constructs assessed have different meaning across groups and scores are therefore not comparable. As noted above, measurement invariance is typically examined at multiple levels. Putnick and Bornstein (2016) described configural invariance as the degree to which the structure is consistent across groups. Metric invariance indicates the degree to which each indicator equally contributes to the latent construct (i.e., positive or negative affect) across groups. Scalar invariance describes the degree to which mean levels of each indicator are equivalent across groups. Lastly, strict invariance describes the degree to which residual variances of each item are equivalent across groups. Establishing measurement invariance of translated measures is important, because it can allow for meaningful cross-cultural research as well as potential cross-cultural comparisons of PANAS scores.

To date, few studies have examined measurement invariance of the PANAS across different cultures and languages. In one of the few studies that has, Lee et al. (2019) assessed measurement invariance of the original English version of the PANAS across American and Singaporean groups and found metric invariance for fifteen of the twenty PANAS items and scalar invariance for only four of the twenty PANAS items (*enthusiastic, scared, nervous, and afraid*). These results suggest that there are differences in mean levels of the remaining sixteen items across groups. Lee et al. (2019) claimed that these findings may be due to individualistic and collectivistic cultural differences between the two groups and warned against direct comparisons of PANAS scores across Asian and Western samples. To date, most researchers have compared Asian countries and Western countries when investigating this topic, with affect receiving relatively little research attention in Arab cultures—which are generally oriented to collectivism. This gap in the literature is likely, at least in part, due to a lack of validated measures of affect for Arabic-speaking groups.

Current Study

In service of establishing the cross-cultural generalizability of PANAS-assessed positive and negative affect, the major aim of the current study was to examine measurement invariance of the English and adapted Arabic versions of the PANAS across two relatively large American and Arab groups of university students. A series of multigroup confirmatory factor analyses (CFA) were fitted to determine the extent to which the two-factor structure, loadings, intercepts, and residual variances are equivalent across American and Arab samples. Establishment of measurement invariance is critical to allow for meaningful cross-cultural research to ensure that any cross-cultural differences are a result of true variation between groups rather than measurement error. In addition to allowing for cross-cultural comparisons of positive and negative affect, an adapted Arabic version of the PANAS may help improve our understanding of various mood-based processes and pathways to mental health, ultimately improving care among Arabic-speaking individuals.

Methods

Participants

Participants included 2449 university students, 979 from the United States (64.2% female) and 1470 (50.0% female) from four Arab nations, described below.

American Group The American group was comprised of two samples of university students in the United States. The first sample included 382 students from a large public university in the southeastern United States ($M_{\text{age}} = 20.5$ years, $SD = 4.5$ years, 52.9% female; Megreya et al. 2018) and the second sample included 597 students from a large public university in the midwestern United States (72.0% female; age data were not collected but all participants were undergraduates enrolled in introductory classes and estimated to be 18–19 years of age). In the first sample, 40.0% of the participants identified as Black/African-American, 29.58% as White, 19.16% as Asian/Pacific Islander, and the remainder identifying as another race/ethnicity. In the second sample, the vast majority of participants identified as White (89.5%); 2.0% identified as Black/African-American, and 4.3% as Asian (4.3%) with the remainder identifying as another race/ethnicity. All procedures were approved by the relevant institutional review boards. Data from the first sample were collected through an online survey while data in the second sample were collected in-person through paper-and-pencil reporting. As a result of missing data, sample sizes ranged from 974 to 979 for individual PANAS items.

Arab Group The Arab group included a sample of 355 students from Egypt ($M_{\text{age}} = 18.6$ years, $SD = 1.4$ years, 50.7% female), a sample of 400 students from Kuwait ($M_{\text{age}} = 18.3$ years, $SD = 1.6$ years, 50% female), a sample of 315 students from the Kingdom of Saudi Arabia ($M_{\text{age}} = 18.5$ years, $SD = 0.9$ years, 49.2% female), and a sample of 400 students from Qatar ($M_{\text{age}} = 18.9$ years, $SD = 1.2$ years, 50% female). All procedures were approved by the relevant institutional review boards and all data were collected in-person via paper-and-pencil surveys. As a result of missing data, sample sizes ranged from 1467 to 1470 for individual PANAS items. Adapted Arabic PANAS data for this group were reported on previously by Megreya et al. (2016).

Measures

Positive and Negative Affect Schedule (PANAS; Watson et al. 1988a, b) The PANAS is a self-report questionnaire, consisting of twenty adjectives which measure positive and negative affect. The current research utilized the trait version of the instrument which asks participants to rate each adjective on a 5-point Likert scale (1 = *very slightly or not at all* to 5 = *extremely*), based on how often they generally experience the emotion. Positive and negative affect scales of both the adapted Arabic version and the English version of the PANAS demonstrated high internal reliability, with Cronbach's alpha ranging from 0.81 to 0.85 for positive affect and 0.80 to 0.84 for negative affect in the Arab group (see Megreya et al. 2018) and 0.90 to 0.92 for positive affect and 0.88 to 0.92 for negative affect in the American group. As previously described by Megreya et al. (2016), in an attempt to capture the essence and meaning of items, the PANAS was translated to Arabic using the committee translation method. The committee translation method involves multiple translators completing a forward translation of the

measure, comparing with each other and with the original version, and then composing a final translated measure. As compared to the popular back-translation method, the committee translation method offers the advantage of multiple members being able to confer and collectively agree on the translation that most closely reflect the meaning of a word in its original language (van de Vijver and Leung 1997). The adapted Arabic language PANAS can be found here: <https://osf.io/95djr/>.

Procedures

The American group consisted of two samples of undergraduate students who participated in research in partial fulfillment of course requirements; one group completed the PANAS online and the other completed the PANAS via an in-person paper-and-pencil survey. For the Arab group, the PANAS was completed using identical procedures across the four countries. Data were collected in-person and in small groups (which ranged from 15 to 35 participants) while they were attending their university classes (for more details, see Megreya et al. 2018). All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from all patients for being included in the study.

Data Analytic Plan

After first examining descriptive statistics, the current study fitted a series of 2-factor CFA models, including a negative affect and a positive affect factor, within a measurement invariance framework to assess whether responses on the PANAS are comparable across American and Arab groups. Consistent with recommendations for CFA-based models for the number of response categories for each item (Raykov 2012; Rigdon 1998), the robust maximum likelihood (MLR) estimator was used for all analyses; all analyses were conducted using Mplus version 8.4 (Muthen and Muthen 1998-2017).

Measurement Invariance As described by Putnick and Bornstein (2016), measurement invariance analyses were conducted by running increasingly restrictive CFA models (i.e., configural, metric, scalar, and strict) to determine whether scores on the PANAS are comparable across American and Arab groups. The American group was used as the reference group for all analyses. Although initially considered, a clustering variable was not used as doing so has been shown to result in biased estimates in models with fewer than 10 clusters (McNeish and Stapleton 2016).

In the configural model, latent means were fixed to zero and variances were fixed to 1, while loadings, intercepts, and residual variances were free. The configural model is the baseline model, which tests whether the structure is consistent across groups. In all subsequent models, latent variances were fixed to one and the first loading was free in the reference group. In the metric model, factor loadings were constrained to equality. The metric model tests the degree to which each observed variable or indicator equally contributes to the latent construct across groups. In the third model, the scalar model, factor loadings, and intercepts were constrained to equality. The scalar model tests for equality of the mean of each indicator across groups. Finally, in the strict model, factor loadings, intercepts, and residual variances

were constrained to equality. The strict model tests for equality of the residual variance of each indicator across groups. Strict invariance is rarely achieved in practice and has been previously described as “overly restrictive” (Little et al. 2007, p. 125). Thus, many researchers argue that scalar—not strict—invariance is the final necessary step (e.g., Milfont and Fischer 2010).

For each model, noninvariance was indicated by a decrease in CFI of more than 0.01 as recommended by Cheung and Rensvold (2002). Additionally, root mean square error of approximation (RMSEA) and standardized root mean square residual (SRMR) were generated and considered. Chi-square difference testing was not used to determine significant differences in model fit as chi-square is overly sensitive to large sample sizes (Lin et al. 2013). If a model was not fully invariant, partial invariance was tested to determine invariance at the individual item level. Partial invariance was tested by iteratively constraining relevant parameters (i.e., loadings, intercepts, or residual variances), one at a time, until model fit declined significantly (i.e., decrease in CFI of more than 0.01). Constraints for items that did not cause a significant decline in model fit were applied to the subsequent model.

Results

Preliminary Analyses

Table 1 shows descriptive statistics for the PANAS in both American and Arab groups. There were significant differences in means for all items with the exception of *hostile* and *inspired*.

Table 1 PANAS descriptive statistics

	American		Arab		Difference in means (99% CI)	<i>t</i>	<i>df</i>	<i>p</i>
	M	SD	M	SD				
Interested	3.421	0.975	3.831	1.090	-0.410 (-0.519, -0.301)	-9.72	2241.5	< 0.0001
Distressed	2.215	1.106	2.452	1.089	-0.237 (-0.353, -0.120)	-5.24	2443.0	< 0.0001
Excited	3.348	1.062	3.593	0.989	-0.245 (-0.354, -0.137)	-5.83	2445.0	< 0.0001
Upset	2.294	1.089	2.550	1.104	-0.256 (-0.373, -0.139)	-5.65	2443.0	< 0.0001
Strong	3.274	1.132	3.522	1.003	-0.248 (-0.364, -0.137)	-5.57	1919.6	< 0.0001
Guilty	1.889	1.070	2.398	1.158	-0.509 (-0.627, -0.391)	-11.14	2193.4	< 0.0001
Scared	1.925	1.056	2.133	1.193	-0.208 (-0.326, -0.089)	-4.53	2260.7	< 0.0001
Hostile	1.649	0.954	1.716	1.078	-0.067 (-0.173, -0.040)	-1.61	2259.9	0.107
Enthusiastic	3.255	1.126	3.717	1.048	-0.462 (-0.576, -0.347)	-10.36	2446.0	< 0.0001
Proud	3.131	1.131	3.898	1.073	-0.767 (-0.883, -0.650)	-16.93	2442.0	< 0.0001
Irritable	2.385	1.105	3.057	1.312	-0.672 (-0.799, -0.546)	-13.67	2311.8	< 0.0001
Alert	3.106	1.033	3.446	1.026	-0.340 (-0.449, -0.230)	-8.00	2447.0	< 0.0001
Ashamed	1.817	1.003	3.396	1.229	-1.579 (-1.695, -1.462)	-34.81	2347.9	< 0.0001
Inspired	3.161	1.172	3.154	1.039	0.007 (-0.112, 0.127)	0.16	1919.7	0.870
Nervous	2.537	1.125	2.978	1.318	-0.441 (-0.569, -0.312)	-8.85	2299.9	< 0.0001
Determined	3.514	1.069	3.359	1.135	0.155 (0.037, 0.273)	3.39	2442.0	0.001
Attentive	3.242	1.022	3.564	0.988	-0.322 (-0.428, -0.215)	-7.79	2447.0	< 0.0001
Jittery	1.863	1.048	2.724	1.398	-0.861 (-0.989, -0.733)	-17.37	2407.6	< 0.0001
Active	3.300	1.097	3.614	0.996	-0.314 (-0.426, -0.201)	-7.18	1954.6	< 0.0001
Afraid	1.951	1.074	2.346	1.222	-0.395 (-0.516, -0.275)	-8.44	2268.7	< 0.0001

*N*American = 974–979, *N*Arab = 1467–1470, as described in the text

Measurement Invariance

Table 2 displays model fit indices for all measurement invariance models at various levels starting with *configural* invariance. At the configural level, the same pattern of fixed and freed parameters was applied and, while fit indices did not meet typically defined cut points, the resulting model evidenced the best fit to the data (CFI=0.821), suggesting that both groups (i.e., American and Arab) share the same factor structure.

Next, we tested measurement invariance at the *metric* level by constraining all factor loadings to be equal across groups. These constraints resulted in a decrease in CFI of more than 0.01 (.023), providing evidence against full metric invariance. Therefore, *partial metric* invariance was tested next. To test for partial metric invariance, factor loadings were iteratively constrained one by one, starting with the item with the smallest difference in loadings across groups, until model fit declined significantly. Partial metric invariance was established for all but the following items: *irritable*, *nervous*, *scared*, and *jittery*. That is, when these four items were constrained to be equal across groups, the CFI decreased by more than 0.01. The remaining 16 items equally contributed to the corresponding factor (positive or negative affect) across groups.

Given evidence for partial metric invariance, we next tested for *scalar* invariance of the sixteen items with metric invariance. To assess for scalar invariance of the sixteen items with metric invariance, loadings and intercepts of these sixteen items were constrained to equality across groups, while loadings and intercepts of *irritable*, *nervous*, *scared*, and *jittery* were free. This resulted in a CFI decrease of more than 0.01 (.068), indicating that scalar invariance of all sixteen items was not supported. Therefore, *partial scalar* invariance was examined by iteratively constraining intercepts of the sixteen items with metric invariance one by one, starting with the item with the smallest difference in intercepts across groups, until model fit declined significantly. Results provided evidence for partial scalar invariance for the following ten items: *active*, *attentive*, *strong*, *inspired*, *excited*, *distressed*, *determined*, *hostile*, *alert*, and *upset*. This suggests that the means of these items are equivalent in both groups.

Table 2 Fit indices for measurement invariance models

	CFI	Chi-square	RMSEA (90% CI)	SRMR
Configural	0.821	3068.773	0.081 (0.079, 0.084)	0.090
Metric	0.798	3442.663	0.084 (0.082, 0.087)	0.074
Partial metric*	0.815	3174.461	0.081 (0.078, 0.084)	0.093
Scalar	0.747	4229.766	0.093 (0.090, 0.095)	0.111
Partial scalar**	0.805	3340.674	0.082 (0.080, 0.085)	0.093
Strict	0.790	3573.802	0.084 (0.082, 0.087)	0.096
Partial strict***	0.795	3500.238	0.083 (0.081, 0.086)	0.096

CFI = comparative fit index. RMSEA = root mean square error of approximation. SRMR = standardized root mean square residual. Best fitting, most restrictive model, shown in **boldface**

*Loadings of the following items are constrained to equality: *attentive*, *alert*, *interested*, *active*, *hostile*, *determined*, *enthusiastic*, *proud*, *strong*, *excited*, *distressed*, *ashamed*, *upset*, *inspired*, *afraid*, *guilty*

**Constraints from the partial metric model apply, intercepts for the following items are constrained to equality: *inspired*, *hostile*, *determined*, *distressed*, *excited*, *strong*, *upset*, *active*, *attentive*, *alert*

***Constraints from the partial scalar model apply, residual variances for the following items are constrained to equality: *active*, *alert*, *attentive*, *strong*, *inspired*, *excited*, *distressed*, *determined*, *hostile*

Given support for partial scalar invariance, we then tested for *strict* invariance of the ten items for which scalar invariance had been established. In the strict model, loadings of the sixteen items with metric invariance and intercepts and residual variances of the ten items with scalar invariance were constrained to be equal across groups, while all other loadings, intercepts, and residual variances were free. The CFI decreased by more than 0.01 (.015) when residual variances of the ten items with scalar invariance were constrained, so *partial strict* invariance was tested by iteratively constraining residual variances of these ten items one by one, starting with the item with the smallest difference in residual variance across groups, until model fit declined significantly. Results provided support for partial strict invariance for all items for which partial scalar invariance was established, excluding *upset*. This indicates that the residual variances of these nine items (i.e., *active, attentive, strong, inspired, excited, distressed, determined, hostile, and alert*) are equivalent across groups.

Therefore, there is evidence to support partial measurement invariance of the PANAS in American and Arab groups. Since the fit of the partial scalar and partial strict models was significantly worse than that of the configural model, the partial metric model was determined to be the best fitting, most restrictive model. Figure 1 shows the partial metric model in the American and Arab groups.

Discussion

The current study examined measurement invariance of the English and adapted Arabic versions of the PANAS in American and Arab university student samples. Using a series of multigroup CFA models, the current study assessed invariance at multiple levels, including full and partial metric, scalar, and strict invariance. Overall, results revealed that the PANAS is partially invariant in American and Arab groups with the partial metric model being the best fitting, most restrictive model.

Configural invariance was established first, suggesting that both American and Arab groups share the same two-factor structure. Next, we tested for full metric invariance and results did not support invariance at this level. Lack of *full* metric invariance indicates that not all factor loadings were equivalent across the two groups. As such, we proceeded to test for *partial* metric invariance. Partial metric invariance analyses revealed that whereas positive affect loadings were equivalent across groups, loadings for four negative affect items varied. Specifically, *irritable, nervous, scared, and jittery* did not have equivalent factor loadings on negative affect across groups, with *irritable, nervous, and jittery* loading more strongly in the Arab group compared to the American group. *Scared*, on the other hand, loaded more strongly on the negative affect factor in the American group compared to the Arab group. This suggests that *irritable, nervous, and jittery* are more representative of negative affect in the Arab group and *scared* is more representative of negative affect in the American group. There are likely a number of potential explanations for this finding. For example, it may be that Arab participants view *irritable, nervous, and jittery* more negatively because of the physiological connotations of these adjectives. On the other hand, *scared* may be interpreted as a short-term experience in response to a specific stimulus.

Consistent with this potential explanation, Thompson (2007) performed a qualitative evaluation of the twenty PANAS items with a focus group of international students at an English-based university in Japan and found that the term *scared* was considered to be “a short-term state in reaction to a given cause” by nonnative English speakers. Interestingly,

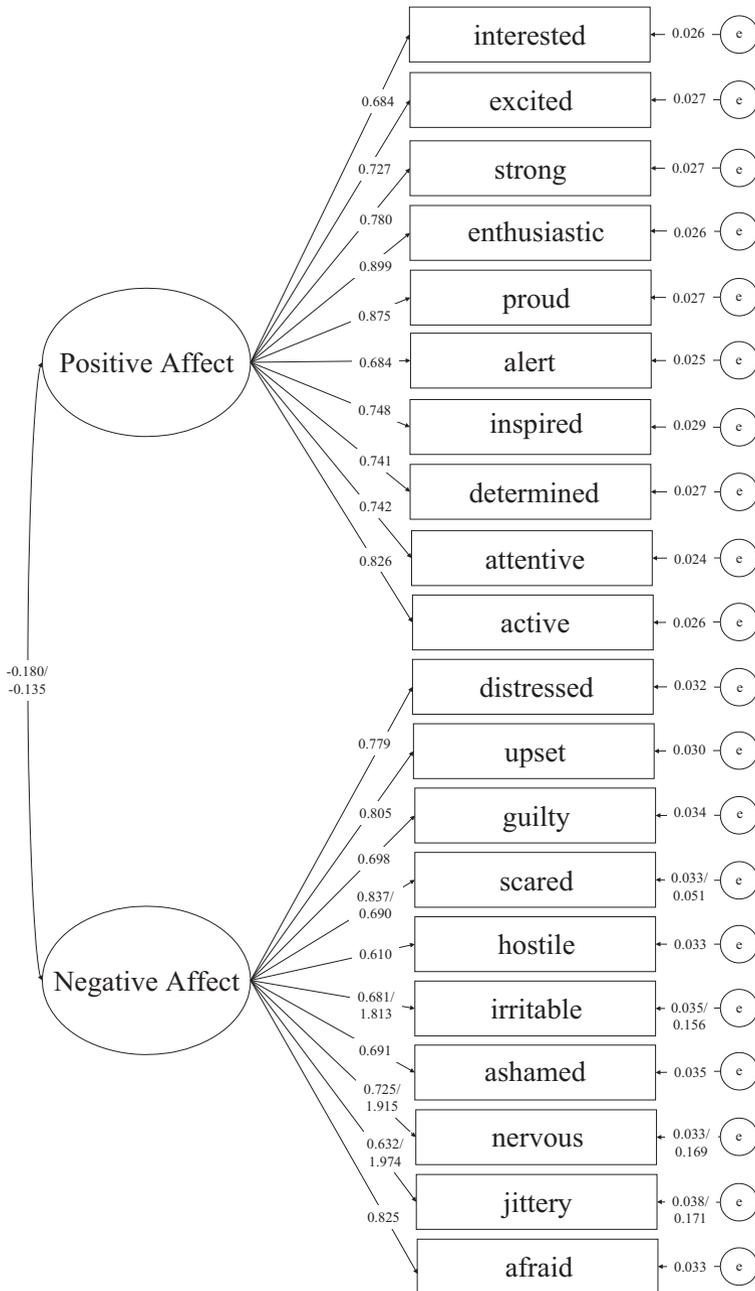


Fig. 1 Partial metric model. *Note:* Loadings (American/Arab) for *scared*, *irritable*, *nervous*, and *jittery* were free, while all other loadings are equal across groups

results did not support metric invariance of *scared*, but did support metric invariance of *afraid*, a word with a very similar meaning in English. This suggests that the adjectives *scared* and *afraid* may be interpreted differently by Arab participants. Given the nuanced difference in

meaning between these two words in Arabic, it is likely that the degree of emotion signified by these adjectives in Arabic may have been interpreted differently than in English with *scared* as more intense than *afraid*.

Our findings deviate from previous research that has established metric invariance of some of these same items in other groups. For example, Lee et al. (2019) found metric invariance of *irritable*, *nervous*, *scared*, and *jittery* in American and Singaporean groups. Since American culture is oriented towards individualism while Singaporean and Arab cultures are both oriented towards collectivism, this suggests that there may be something unique about negative affect in Arab culture, above and beyond collectivistic culture. Few researchers have studied cross-cultural differences in negative affect using an Arab sample, so it is unclear whether or not there are specific differences in the conceptualization of negative affect among those in Arab countries and those in other collectivistic countries, such as China and Japan.

Evidence supported partial metric invariance, so we proceeded to test for scalar invariance of the sixteen items for which metric invariance had been established. We did not find evidence to support scalar invariance of all sixteen items, so we then tested for partial scalar invariance. Scalar noninvariance of *ashamed*, *guilty*, *proud*, *enthusiastic*, *afraid*, and *interested* suggests that the mean levels of these items are different between the two groups, with intercepts for these items being greater in the Arab group compared to the American group. This could be due to individualistic and collectivistic cultural differences between the two groups, but as previously stated, few researchers have investigated cross-cultural differences in affect with an Arab sample. In addition, translation of the PANAS from English to Arabic may have led to differences in the interpretation of noninvariant items between the American and Arab groups.

Limitations and Future Directions

The current study is not without limitations. Whereas the current study was not focused specifically on attempting to estimate the best fitting model in absolute model fit terms, it is important to note that the model fits were lower than expected. Given its widespread use and translation into multiple languages, it will be important for future research to clarify an optimally fitting structural model of the adapted Arabic version of the PANAS. Additionally, although the PANAS was translated in a manner designed to optimize the essence and meaning of adjectives (Megreya et al. 2016), there are often indirect translations or multiple translations of words across languages. Further, we assessed measurement invariance of the English and adapted Arabic versions of the PANAS with participants from four Arabic-speaking countries (i.e., Egypt, Kuwait, Qatar, and the Kingdom of Saudi Arabia) and only one English-speaking country (i.e., the United States). We considered including a clustering variable to account for this, but given the relatively small number of potential clusters, the use of a clustering variable in this circumstance is not recommended (McNeish and Stapleton 2016). It will thus be important for future studies to consider invariance between Arabic speaking countries. Whereas the current research included participants from four Arab countries, future studies that examine measurement invariance of the adapted Arabic version of the PANAS are needed that include Arabic-speaking participants from countries not included here. Further, it might be wise for future studies to more formally consult with experts in the Arabic language about potential interpretation differences of the four items evidencing metric noninvariance. As all of our participants were university students, future studies should examine measurement invariance of the adapted Arabic version of the PANAS across age

and education levels as well as within more clinical samples. Finally, researchers should make future efforts to further understand differences in the experience and expression of affect between individualistic and collectivistic cultures, including Arab samples.

Conclusions

The current study aimed to assess measurement invariance of the English and adapted Arabic versions of the PANAS in American and Arab university student samples. Multigroup confirmatory factor analysis revealed that the PANAS is partially invariant in American and Arab university students. Whereas all ten positive affect items evidenced metric invariance, 4 negative affect items, *irritable*, *nervous*, *scared*, and *jittery*, demonstrated nonequivalent loadings (metric noninvariance) on the negative affect factor in American and Arab groups.

Overall, as sixteen of the twenty PANAS items evidenced metric invariance, the adapted Arabic PANAS appears to be generally appropriate for use in the Arab world. Indeed, all items on the positive affect scale evidenced metric invariance and can readily be used to make meaningful cross-cultural comparisons. Items on the negative affect scale less consistently met criteria for metric invariance with *irritable*, *nervous*, *scared*, and *jittery*, failing to evidence metric invariance. With that important caveat in mind, it will be imperative for future research to more explicitly consider the underlying causes of metric noninvariance of these four items and how measurement of PANAS-based negative affect can be improved, if needed. Indeed, replication of this study with subjects from a wider variety of Arab nations will be important in order to better understand measurement invariance of the English and adapted Arabic versions of the PANAS. To date, there are few Arabic language measures of positive and negative affect. Thus, despite the four noninvariant items that will require additional research to interpret, it is recommended that researchers continue to use and test the adapted Arabic version of the PANAS. The Arabic adapted PANAS is likely to continue to prove useful in more fully explicating affect variation in the Arab world and improving our understanding of various mood-based pathways to psychopathological outcomes in this region. Future research using this version will also allow for potential refinement of the instrument if and as needed.

Compliance with Ethical Standards

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

Ethical Approval All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000.

Informed Consent Informed consent was obtained from all patients for being included in the study.

Appendix 1

يتكون هذا المقياس من عدد من الكلمات التي قد تصف مشاعر أو انفعالات مختلفة. أقرأ كل كلمة ثم سجل إجابتك في الفراغ الذي يجاورها. المطلوب منك تحديد إلى أي مدى قد تشعر بوجه عام بهذه الطريقة، بمعنى آخر حاول تحديد إلى أي مدى قد تصفك هذه الصفات بوجه عام. استخدم مسطرة الدرجات التالية في تسجيل إجابتك.

5 كثير جدا	4 كثير إلى حد ما	3 متوسط	2 قليل	1 قليل جدا أو ليس على الإطلاق
___	سريع الغضب		___	مُهْتَم
___	يقظ		___	حزين
___	خجول		___	فرحان
___	مُلْهَم		___	متضايق
___	عصبي		___	قوي
___	حازم		___	مُذنب
___	مُنْتَبِه		___	مرعوب
___	شديد العصبية		___	عدائي
___	نشط		___	مُتحمس
___	خائف		___	فخور

References

- Bodenhausen, G. V., Sheppard, L. A., & Kramer, G. P. (1994). Negative affect and social judgment: The differential impact of anger and sadness. *European Journal of Social Psychology, 24*(1), 45–62. <https://doi.org/10.1002/ejsp.2420240104>.
- Bryan, T., Sullivan-Burstein, K., & Mathur, S. (1998). The influence of affect on social information processing. *Journal of Learning Difficulties, 31*, 418–426. <https://doi.org/10.1177/002221949803100501>.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5.
- Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 43*(3), 245–265. <https://doi.org/10.1348/0144665031752934>.
- Díaz-García, A., González-Robles, A., Mor, S., Mira, A., Quero, S., García-Palacios, A., Baños, R. M., & Botella, C. (2020). Positive and Negative Affect Schedule (PANAS): Psychometric properties of the online

- Spanish version in a clinical sample with emotional disorders. *BMC Psychiatry*, 20(1), 56. <https://doi.org/10.1186/s12888-020-2472-1>.
- Dockray, S., & Steptoe, A. (2010). Positive affect and psychobiological processes. *Neuroscience and Biobehavioral Reviews*, 35(1), 69–75. <https://doi.org/10.1016/j.neubiorev.2010.01.006>.
- Estévez-López, F., Pulido-Martos, M., Armitage, C. J., Wearden, A., Álvarez-Gallardo, I. C., Arrayás-Grajera, M. J., Girela-Rejón, M. J., Carbonell-Baeza, A., Aparicio, V. A., Geenen, R., Delgado-Fernández, M., & Segura-Jiménez, V. (2016). Factor structure of the Positive and Negative Affect Schedule (PANAS) in adult women with fibromyalgia from Southern Spain: The al-Ándalus project. *PeerJ*, 4, e1822. <https://doi.org/10.7717/peerj.1822>.
- Joiner Jr., T. E., Sandin, B., Chorot, P., Lostao, L., & Marquina, G. (1997). Development and factor analytic validation of the SPANAS among women in Spain: (More) Cross-cultural convergence in the structure of mood. *Journal of Personality Assessment*, 68(3), 600–615. https://doi.org/10.1207/s15327752jpa6803_8.
- Lee, S. T. H., Hartanto, A., Yong, J. C., Koh, B., & Leung, A. K. (2019). Examining the cross-cultural validity of the positive affect and negative affect schedule between an Asian (Singaporean) sample and a Western (American) sample. *Asian Journal of Social Psychology*, 23, 109–116. <https://doi.org/10.1111/ajsp.12390>.
- Lin, M., Lucas Jr., H. C., & Shmueli, G. (2013). Too big to fail: Large samples and the p-value problem. *Information Systems Research*, 24(4), 906–917. <https://doi.org/10.1287/isre.2013.0480>.
- Little, T. D., Card, N. A., Slegers, D. W., & Ledford, E. C. (2007). Representing contextual effects in multiple-group MACS models. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (p. 121–147). Lawrence Erlbaum Associates Publishers.
- McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review*, 28(2), 295–314. <https://doi.org/10.1007/s10648-014-9287-x>.
- Megreya, A. M., Latzman, R. D., Al-Attayah, A. A., & Alrashidi, M. (2016). The robustness of the nine-factor structure of the Cognitive Emotion Regulation Questionnaire across four Arabic-speaking Middle Eastern countries. *Journal of Cross-Cultural Psychology*, 47(6), 875–890. <https://doi.org/10.1177/0022022116644785>.
- Megreya, A. M., Latzman, R. D., Al-Emadi, A. A., & Al-Attayah, A. A. (2018). An integrative model of emotion regulation and associations with positive and negative affectivity across four Arabic speaking countries and the USA. *Motivation and Emotion*, 42(4), 566–575. <https://doi.org/10.1007/s11031-018-9682-6>.
- Merz, E. L., Malcarne, V. L., Roesch, S. C., Ko, C. M., Emerson, M., Roma, V. G., & Sadler, G. R. (2013). Psychometric properties of Positive and Negative Affect Schedule (PANAS) original and short forms in an African American community sample. *Journal of Affective Disorders*, 151(3), 942–949. <https://doi.org/10.1016/j.jad.2013.08.011>.
- Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3(1), 111–130. <https://doi.org/10.21500/20112084.857>.
- Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide. Eighth Edition*. Los Angeles: Muthén & Muthén.
- Pandey, R., & Shrivastava, N. (2008). Psychometric evaluation of a Hindi version of Positive and Negative Affect Schedule. *Industrial Psychiatry Journal*, 17(1), 49–54.
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>.
- Raykov, T. (2012). Scale construction and development using structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 472–492). New York, NY: The Guilford Press.
- Rigdon, E. E. (1998). Structural equation modeling. In G. A. Marcoulides (Ed.), *Methodology for business and management. Modern methods for business research* (p. 251–294). Lawrence Erlbaum Associates Publishers.
- Serafini, K., Malin-Mayor, B., Nich, C., Hunkele, K., & Carroll, K. M. (2016). Psychometric properties of the Positive and Negative Affect Schedule (PANAS) in a heterogeneous sample of substance users. *The American Journal of Drug and Alcohol Abuse*, 42(2), 203–212. <https://doi.org/10.3109/00952990.2015.1133632>.
- Terracciano, A., McCrae, R. R., & Costa Jr., P. T. (2003). Factorial and construct validity of the Italian Positive and Negative Affect Schedule (PANAS). *European Journal of Psychological Assessment*, 19(2), 131–141. <https://doi.org/10.1027/1015-5759.19.2.131>.
- Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology*, 38(2), 227–242. <https://doi.org/10.1177/0022022106297301>.

- Tuccitto, D. E., Giacobbi, P. R., & Leite, W. L. (2010). The internal structure of positive and negative affect: A confirmatory factor analysis of the PANAS. *Educational and Psychological Measurement, 70*(1), 125–141. <https://doi.org/10.1177/0013164409344522>.
- van de Vijver, F. J. R., & Leung, K. (1997). Methods and data analysis of comparative research. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology* (2nd ed., pp. 257–300). Boston: Allyn & Bacon.
- Watson, D., Clark, L. A., & Carey, G. (1988a). Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology, 97*(3), 346–353. <https://doi.org/10.1037/0021-843X.97.3.346>.
- Watson, D., Clark, L. A., & Tellegen, A. (1988b). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. <https://doi.org/10.1037/0022-3514.54.6.1063>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.