On the Structure of Affect: A Meta-Analytic Investigation of the Dimensionality and the Cross-National Applicability of the Positive and Negative Affect Schedule (PANAS)

Nadine Wedderhoff [1], Timo Gnambs [2, 3], Oliver Wedderhoff [4], Tanja Burgard[4] & Michael Bosnjak [1, 4]

[1] Department of Psychology, University of Trier, Trier, Germany

[2] Institute for Education and Psychology, University of Linz, Linz, Austria

[3] Leibniz Institute for Educational Trajectories, Bamberg, Germany

[4] Leibniz Institute for Psychology Information (ZPID), Trier, Germany

Address for correspondence:

Nadine Wedderhoff, Department of Psychology, University of Trier, Universitätsring 15, 54296 Trier, Germany. Email: kasten@uni-trier.de. Tel: +49 651 201 2738

Accepted for publication at the *Zeitschrift für Psychologie*.

Acknowledgments

In this paper, we make use of the LISS (Longitudinal Internet Studies for the Social Sciences) panel administered by CentERdata (Tilburg University, The Netherlands). Additionally, we make use of data that were collected by the Laboratory for the Study of Stress, Immunity, and Disease at Carnegie Mellon University under the directorship of Sheldon Cohen; and were accessed via the Common Cold Project (CCP) Website (www.commoncoldproject.com). CCP data are made publicly available through a grant from the National Center for Complementary and Integrative Health (AT006694); the conduct of the studies was supported by grants from the National Institute of Allergy and Infectious Diseases (AI066367); and secondary support was provided by a grant from the National Institutes of Health to the University of Pittsburgh Clinical and Translational Science Institute (RR024153).

Further data was from the public release of the German Survey on Volunteering (FWS), provided by the Research Data Centre of the German Centre of Gerontology (FDZ-DZA).

Abstract

The *Positive and Negative Affect Schedule* (PANAS; Watson, Clark, & Tellegen, 1988) is a popular self-report questionnaire that is administered all over the world. Though originally developed to measure two independent factors, different models have been proposed in the literature. Comparisons among alternative models as well as analyses concerning their robustness in cross-national research have left an inconclusive picture. Therefore, the present study evaluates the dimensionality of the PANAS and differences between English and translated versions of the PANAS using a meta-analytic structural equation modeling approach. Correlation matrices from 57 independent samples (N = 54,043) were pooled across subsamples. For both English and non-English samples, a correlated two-factor model including correlated uniquenesses provided the best fit. However, measurement invariance analyses indicated differences in factor loadings between subsamples. Thus, cross-national application of the PANAS might only be justified if measurement equivalence was explicitly tested for the countries at hand.

Keywords: PANAS, Positive and Negative Affect Schedule, factor analysis, meta-analytic structural equation modeling, measurement invariance

On the Structure of Affect: A Meta-Analytic Investigation of the Dimensionality and the Cross-National Applicability of the Positive and Negative Affect Schedule (PANAS)

The 20-item Positive and Negative Affect Schedule (PANAS; Watson et al., 1988) is a widely applied self-report measure to assess two broad domains of affect, namely *Positive Affect* (PA) and *Negative Affect* (NA). The former is associated with pleasurable engagement with the environment, whereas the latter reflects a dimension of general distress summarizing a variety of negative states such as anger, guilt, or anxiety. Items are simply mood-related adjectives and participants are asked to rate the extent to which they have experienced the particular emotion with reference to a given time frame on a 5-point scale ranging from 1 (*very slightly or not at all*) to 5 (*very much*). The PANAS is suitable for both measuring state and trait affectivity, depending on which specific instruction is applied. Its brevity and frequently demonstrated good psychometric properties (e.g., Crawford & Henry, 2004; Leue & Lange, 2011) have contributed to the widespread use of the PANAS in many areas of psychology. Over the years modified versions of the PANAS were developed to meet the requirements of certain subpopulations or application contexts, including a version suitable for children (PANAS-C; Laurent et al., 1999), different short-forms (PANAS-SF and I-PANAS-SF; Kercher, 1992; Thompson, 2007), as well as an expanded form consisting of 60 items (PANAS-X; Watson & Clark, 1994).

However, despite its popularity, there is still an ongoing discussion concerning its internal structure. Specifically, since its development, there is disagreement about both the underlying dimensionality of the PANAS as well as the interrelatedness of PA and NA and previous studies that evaluated the internal structure (e.g., Estévez-López et al., 2016; Leue & Beauducel, 2011; Vera-Villarroel et al., 2019) produced inconsistent results. Therefore, this study tends to evaluate the internal structure of the PANAS using a meta-analytic approach.

The Dimensionality of the PANAS

Although the terms positive (PA) and negative affect (NA) might suggest that these two are on opposite ends of a bipolar affect scale, Watson and colleagues (Watson et al., 1988; Watson & Clark, 1997; Watson & Tellegen, 1985; Watson, Wiese, Vaidya, & Tellegen, 1999) conceptualized PA and NA as distinctive and independent dimensions of affective experience. In this vein, it is expected to be possible to experience, for instance, high positive and high negative affect at the same time. Accordingly, the goal underlying the development of the PANAS was twofold: (1) To provide brief and independent measures of PA and NA that are most appropriately represented by an orthogonal two-factor model and (2) to cover a broad and representative range of moods. For this reason, Watson et al. (1988) based the development of the PANAS items on the work of Zevon and Tellegen (1982) who identified 60 mood descriptors that could be assigned to 20 content categories, which in turn were associated with either positive or negative affect. From those 60 items, Watson et al. (1988) chose adjectives that represented relatively pure markers of either PA or NA. That is, they selected those items that exhibited high loadings on one factor with little to no cross-loadings on the other factor. Afterward, the factorial validity was examined using different samples (mostly university students), that rated the frequency of the specific mood terms for six different timeframes (e.g., at the moment, past few days, in general, etc.). In line with the proposed structure, principal factor analyses with varimax rotation revealed two dominant factors in each sample. Table 1 displays all original PANAS items as well as their allocation to the PA and NA dimensions and the content categories according to Zevon and Tellegen (1982).

Since the introduction of the PANAS, many studies examined its factorial validity using exploratory (EFA) or confirmatory factor analysis (CFA) and have indeed come to different conclusions about which measurement model fits best (see Figure 1 for an overview of the models presented in the literature). Although EFA studies using varimax rotation typically suggested the existence of two independent factors (e.g. Krohne, Egloff, Kohlmann, & Tausch, 1996), confirmatory approaches often failed to provide an appropriate model fit for the orthogonal two-factor model (e.g. Crawford & Henry, 2004; Leue & Beauducel, 2011; Terracciano, McCrae, & Costa, 2003; Villodas, Villodas, & Roesch, 2011). Particularly, the orthogonality between PA and NA has frequently been questioned. In this vein, many studies report a significant improvement in model fit when the two factors were allowed to correlate (e.g. Joiner, Sandín, Chorot, Lostao, & Marquina, 1997; Merz et al., 2013). However, the empirical results regarding the fit of a correlated two-factor model are also mixed. In many studies, the correlation between the two factors was low to moderate indicating that they shared at most 9.00 % of the common variance (Tuccitto, Giacobbi, & Leite, 2010). Additionally, in some studies, the oblique two-factor model only provided adequate fit when further modifications were adopted (Crawford & Henry, 2004; Merz & Roesch, 2011; Rush & Hofer, 2014; Tuccitto et al., 2010). For example, Crawford and Henry (2004) administered the PANAS to N = 1,003 adults in the United Kingdom and evaluated competing models of the latent structure using CFA. Both the orthogonal two-factor model as well as the oblique counterpart, which included the intercorrelation between the factors, resulted in poor model fit. However, the latter representation exhibited appropriate fit when residuals were allowed to correlate in accordance to the content categories defined by Zevon and Tellegen (1982). That is, correlated errors were permitted for adjectives that were drawn from the same category.

Some authors argued that these modifications cover the actual structure underlying the PANAS responses and therefore specified more complex models such as hierarchical second-order models (Egloff, Schmukle, Burns, Kohlmann, & Hock, 2003; Mehrabian, 1997; Mihic, Novovic, Colovic, & Smederevac, 2014), or different variants of three-factor models (Beck et al., 2003; Gaudreau, Sanchez, & Blondin, 2006; Killgore, 2000), or bi-factor representations (Leue & Beauducel, 2011; Seib-Pfeifer, Pugnaghi, Beauducel, & Leue, 2017). Although different in detail, many authors that favor complex models over the two-factor representation of the PANAS provided evidence for a differentiation of the NA factor into further subscales. For instance, Killgore (2000) used EFA on PANAS responses from *N* = 302 university students and found three principal components with an eigenvalue over one that accounted for 51.3 % of the variance. Whereas the first factor was congruent with the PA factor according to Watson et al. (1988), the NA items could be assigned to two distinct factors: *Upset* and *afraid*. This solution was notably similar to those found by Mehrabian (1997) who modeled upset and afraid as two lower-order factors within a hierarchical model. However, compared to Killgore (2000) the assignment of the items to the NA factors was slightly different and failed to replicate in other studies (Beck et al., 2003; Gaudreau et al., 2006; Vera-Villarroel et al., 2019). Additionally, correlations between the NA sub-facets are commonly very high (e.g., up to *r* = .85; Heubeck & Wilkinson, 2019), raising concerns about the practical relevance and the need to distinguish between these factors. Similarly, efforts to further differentiate the PA factor into sub-facets also failed (Egloff et al., 2003; Mihic et al., 2014).

Finally, Leue and Lange (2011) applied the PANAS to *N* = 354 adult participants in Germany and introduced a bi-factor representation of its structure. Within this model, each item loads on one of the specific factors, PA or NA, as well as on a general factor that was expected to represent a fundamental approach or withdrawal tendency (i.e., *affective polarity*). This model exhibited superior model fit compared to alternative CFA models including the uncorrelated and the correlated two-factor model as well as the three-factor model proposed by Gaudreau et al. (2006). Additionally, Seib-Pfeifer et al. (2017) replicated the superiority of the bi-factor representation and provided further evidence on the measurement invariance across men and women using multiple-group CFA. However, in a recent study, the bi-factor formulation was harshly criticized (Heubeck & Wilkinson, 2019). Although it provided a good fit to the PANAS responses of *N* = 2,392 Australian adults, the inspection of factor loadings indicated a weak general factor that did not replicate across different samples.

Cross-National Replicability of the PANAS Structure

The PANAS is administered all over the world and has been adapted to various languages, including for instance German (Krohne et al., 1996), Italian (Terracciano et al., 2003), Dutch (Engelen, Peuter, van Diest, & van den Bergh, 2006), Spanish (Joiner et al., 1997), Turkish (Gençöz, 2000), Urdu (Akhter, 2017), and Serbian (Mihic et al., 2014). So far, there has been limited investigation of the equivalence of original and translated versions of the PANAS. Moreover, the evidence that translated versions of the PANAS measure the same constructs as the original questionnaire is often rather assumed than empirically tested. However, the proof of factorial invariance is essential to compare scores in diverse samples, as differences can reflect true differences in the constructs but they can also be rooted in differences in how groups experience the questionnaire items (Milfont & Fischer, 2010) or cultural differences in the susceptibility to response biases, like acquiescence or extreme responding (Kemmelmeier, 2016).

Given the equivocal findings regarding the internal structure of the PANAS as well as the widespread application of different translations, it can be assumed that inconsistencies are to some degree attributable to language differences in test versions. Notable, non-English versions of the PANAS greatly differ in the rigor of the translation procedure leading to potential differences in the meaning of some items. Moreover, given that the goal of a successful translation process is not a word-for-word translation but maintaining the meaning (Sireci, Yang, Harter, & Ehrlich, 2006), the translation of adjective-based questionnaires like the PANAS might be comparable difficult, as “single words do not sufficiently convey the meaning of these traits” (Nye, Roberts, Saucier, & Zhou, 2008; p. 1534). In addition, the translation may be more difficult for some languages than for others. For instance, Mihic et al. (2014) noted that there is no straightforward Serbian word for “distressed”.

In addition to the obstacles associated with the translation process, the emic development approach underlying the PANAS items might put the factorial invariance at risk. That is, items might exhibit ambiguous meanings in some cultures even if translated correctly. For example, Jackson and Chen (2008) applied a translated version of the PANAS to 593 Chinese students. They were not able to replicate the conceptualization of Watson et al. (1988) since the item “alert” significantly correlates with PA and NA, suggesting that this term has an ambiguous meaning. Moreover, further studies reported on item-specific problems with the cross-cultural adaptation of the PANAS (Gaudreau et al., 2006; Mackinnon et al., 1999; Mihic et al., 2014).

The Present Study and Hypotheses

In light of the inconsistent findings regarding the dimensionality of the PANAS scales, the goal of the present study was to provide evidence for the internal structure of the PANAS using a meta-analytical structural equation modeling (MASEM) approach (Cheung & Chan, 2005). We compared the fit of nine different competing measurement models described in the literature. Given the overwhelming evidence for the differentiation of PA and NA (e.g., Crawford & Henry, 2004; Merz et al., 2013), we expected a two-factor model to perform best. However, contrary to the originally proposed structure, PA and NA are not expected to be orthogonal but moderately correlated. Moreover, measurement invariance between English and translated PANAS versions will be tested. Since many studies reported on item-specific problems with the cross-cultural adaptation of the PANAS (e.g. Gaudreau et al., 2006; Jackson & Chen, 2008; Mackinnon et al., 1999; Mihic et al., 2014), it is expected that there are differences between English and translated versions of the PANAS.

Method

Literature Search and Study Selection

The literature search encompassed articles and datasets published between 1988 – the year the PANAS has been published – to May 2020. To locate relevant studies, we conducted electronic searches using major scientific (PsycINFO, ERIC, Psyndex, Medline) as well as non-scientific databases (Google Scholar) using the search term *PANAS* or "*Positive and Negative Affect Schedule"* in combination with "*correlation matrix"*. Because our meta-analytic analysis approach is based on inter-item correlations - which are rarely reported in articles - we extended our search to scientific data repositories including ICPSR data archive, PsychData, UK data archive, GESIS data catalogue, figshare, and the Open Science Framework. Additionally, about 80 corresponding authors were contacted and asked if they would be willing to share their PANAS data or a correlation matrix between the 20 items. Given these search strategies, we found 8,584 potentially eligible studies and data sets. After reviewing the title and abstracts of the research objects, 178 records remained in the selection process, which were then subjected to detailed scrutiny. Studies were included in the meta-analytic data set dependent upon the following eligibility criteria: (1) Studies had to apply the original 20 items PANAS. Longer versions were only considered if all original items were implemented. Short forms of the PANAS were excluded. (2) Studies had to report inter-item correlations between the 20 items. Alternatively, the raw data file had to report item-level responses to the items. A PRISMA flow chart is presented in Figure 2.

Data Extraction and Coding Procedure

Pearson product-moment-correlation coefficients between the 20 PANAS items were used as effect size measures. In one case, the relevant inter-item correlation matrix was printed in the article itself. The remaining articles provided access to raw data that allowed the calculation of the respective correlations.

In addition, a coding scheme was developed to assess descriptive information of the samples and application settings. Accordingly, for each study we coded the following characteristics: (a) the publication year, (b) the country in which the study was conducted, (c) the language in which the PANAS was presented, (d) the sample size, (e) the percentage of female participants, (f) the mean age of the sample, and (g) what kind of instruction was used. All studies were coded by the first author. Twenty studies were additionally coded by another coder to calculate interrater-reliability. Interrater reliability indicated almost perfect agreement with a mean Krippendorff's alpha of .997. All existing discrepancies were solved by discussion.

Statistical Analyses

To compare different factor models of the PANAS we applied the meta-analytic structural equation modeling approach proposed by Cheung and Chan (2005) which is basically a two-stage procedure. In stage 1, correlation coefficients are pooled using a multigroup structural equation modeling approach. In stage 2, the resulting pooled correlation matrix is used to test competing CFA models using a weighted least square estimation. The asymptotic covariance matrix from the first step is used as a weight matrix to ensure "that correlation coefficients that are estimated with more precision (based on more studies) in Stage 1 get more weight in the estimation of model parameter in Stage 2" (Jak, 2015; p.26). To analyze potential moderating effects of language, subgroup analyses were applied as presented by Jak and Cheung (2018a). Thus, we first estimated the same factor structure in each subgroup independently (*configural invariance* model) and, subsequently, constrained the factor loadings across groups (*metric invariance* model). If the latter model did not result in a deteriorated fit, measurement invariance across groups could be established. All computations were conducted in R using the metaSEM package version 1.2.3.1 (Cheung, 2014).

A total of nine different factor structures are tested that have been proposed in the literature for the PANAS. This includes two- and three-factor models as well as two different variants of hierarchical factor structures and a bi-factor representation. For each model, the pattern of item loadings is illustrated in Figure 1.

Two-factor models. Many authors who evaluated the internal structure of the PANAS adopted a two-factor representation of affectivity. Within this study, we tested three different models that build upon the differentiation of positive and negative affect. Model 1a represented the orthogonal two-factor model originally proposed by Watson et al. (1988). Model 1b was identical to model 1a except that Model 1b allows for the intercorrelation between the two latent factors. Finally, Model 1c is identical to 1b but correlated errors were permitted following Zevon and Tellegen's (1982) content categories.

Three-factor models. Three variants of CFA models were included that expressed the research hypotheses that three facets of affect can be found within the PANAS. In this vein, Model 2a is the three-factor specification proposed by Killgore (2000), who differentiated PA (items: 1, 3, 5, 9, 10, 12 ,14, 16, 17, 19), upset (items: 2, 4, 6, 8, 11, 13) and afraid (items: 7, 15, 18, 20). Model 2b was proposed by Gaudreau et al. (2006) and is largely consistent with the representation by Killgore (2000), but item 2 is assigned to the afraid factor and not the upset factor. For both models, latent factors are allowed to correlate. Model 2c is based on Beck et al. (2003) who also categorized negative affect items in two distinctive factors. However, here the factors were labeled anxiety/anger (items: 2, 4, 7, 8, 11, 15, 18, 20) and guilt/shame (items: 6, 13), and latent factors are uncorrelated.

Hierarchical models. Two models were specified that propose a hierarchical structure underlying the PANAS responses. Model 3a represents the hierarchical structure defined by Mihic et al. (2014) who conceptualized PA and NA as second-order factors both consisting of three first-order factors. PA items were categorized in items that reflect joviality (items:1, 3, 5), self-assurance (items: 9, 10, 12), and attentiveness (items: 14, 16, 17 19), whereas NA comprises three factors labeled afraid (items: 2, 4, 6, 7, 8), self-disgust (items: 11, 13, 15), and hostility (items: 18, 20).

Model 3b was the model proposed by Mehrabian (1997) who conceptualized NA as a second-order factor, that comprises the two distinct first-order factors *upset* and *afraid*. The former comprised six items (6, 7, 13, 15, 18, 20), whereas the latter comprised four (2, 4, 8, 11).

Bi-factor model. The bi-factor model specified by Leue and colleagues (Leue & Beauducel, 2011; Seib-Pfeifer et al., 2017) is presented in Model 4a. It includes two specific factors NA and PA as well as a general factor on which all items load. All latent factors were uncorrelated, as this would complicate interpretability.

The evaluation of model fit was based on the comparative fit index (CFI; Bentler, 1990), the standardized root mean square residual (SRMR), and the root mean square error of approximation (RMSEA). Based on common recommendations from Hu & Bentler (1999), models exhibiting CFI > .95, RMSEA ≤ .08; and SRMR ≤ .10 were considered to provide acceptable fit, and CFI > .97, RMSEA ≤ .05; and SRMR ≤ .05 was interpreted as good fit. Additionally, to provide a comparison of model fit, *Akaike's Information Criterion* (AIC) and the *Bayesian Information Criterion* (BIC) were used, with smaller values indicating better fit respectively.

Open Practices

To foster transparency and reproducibility of our research, all data (doi: http://dx.doi.org/10.23668/psycharchives.4272) as well as the corresponding R code (doi: http://dx.doi.org/10.23668/psycharchives.4271) and additional supplementary material (doi:

http://dx.doi.org/10.23668/psycharchives.4274) is openly accessible on PsychArchives (https://www.psycharchives.org/).

Results

Study Characteristics

Overall, we found 47 studies reporting correlation matrices from 57 independent samples to be eligible for the current meta-analysis (see Figure 2). The cumulative sample size was N = 54,043. The average study was based on M = 948 participants (median = 430; SD = 1,533.85, min = 30, max = 6,762), included M = 61 % women and was published in M = 2014 (SD = 4.53 years). Most studies were conducted in the Netherlands (k = 15, 26.3 %), the USA (k = 13, 22.8 %) or Germany (k = 9, 15.8 %). Accordingly, the predominant languages were English (k = 22, 39.3 %), Dutch (k = 15, 26.8 %), and German (k = 9, 16.1 %). Regarding the instruction type, about half of the studies applied a state measure of affect, that is participants were asked about their current mood (k = 26, 45.6 %). A trait measure (i.e., participants indicate how they feel in general) was used in k = 9 (15.8 %) cases, whereas the other studies applied another type of instruction.

Pooled Correlation Matrix

In two matrices, correlations for one item were missing. However, previous research showed that the applied multivariate method performs well in light of missing correlation coefficients (Jak & Cheung, 2018b). Following the procedure described by Cheung and Chan (2005), we pooled the correlation matrices across samples. Given the high number of elements in every correlation matrix (i.e., 190 elements), it is less surprising that a random-effects model did not converge. The fixed-effect model for the overall sample indicated approximate fit according to RMSEA (.068) but not to SRMR (.104). Since heterogeneity of correlation matrices is expected to be attributable to differences in the applied test version (i.e., whether the original or a translated version of the PANAS is used), we are using subgroup analysis right away (Jak & Cheung, 2018a). Pooled correlation matrices are presented Table 2. Overall, 22 independent samples applied the English version of the PANAS (n = 7,487), whereas 35 samples used a translated version (n = 46,556). Correlations between positive affect items were all positive and substantial for both English ($\bar{r}$ = .467, Range: .316 to .634) and non-English samples ($\bar{r}$ = .398, Range: .239 to .574). Similarly, all negative affect items correlated positive with medium to large effect sizes (English samples: $\bar{r}$ = .444, Range: .330 to .743; non-English samples: $\bar{r}$ = .444, Range: .308 to .621). Additionally, discriminant correlations between positive and negative affect items were all negligible with $\bar{r}$ = -.091 (Range: -.156 to -.005) for English items and $\bar{r}$ = .057 (Range: -.150 to .034) for translated items.

Evaluating Model Fit

Table 3 summarizes the fit indices for the competing CFA models separately for English and non-English samples. Regarding the former subsample, the oblique two-factor model including correlated errors according to the content categories proposed by Zevon and Tellegen (1982) exhibited the best fit. Regarding the non-English samples, none of the tested models achieved acceptable fit according to CFI. On the other hand, all models except for the three-factor model according to Beck et al. (2003) exhibited at least appropriate fit according to RMSEA and SRMR. Based on the information criteria, we concluded that the bi-factor specification from Leue and Beauducel (2011) seemingly provided the best model fit for the translated versions. However, a further inspection of factor loadings and explained common variances indicated that this solution is anomalous (Eid, Geiser, Koch, & Heene, 2017). That is, all factor loadings are positive and substantial (all $\lambda \geq .580$) for the specific factor positive affect, whereas negative affect exhibits an inconsistent loading pattern including both positive and negative loadings which are additionally all rather small ($|\lambda| \leq .360$). Correspondingly, the explained common variance (ECV, Rodriguez, Reise, & Haviland, 2016) showed a negligible relative strength of the NA factor (i.e., ECV = .044) indicating that less than 5 % of the variability in responses is due to the specific NA factor. Moreover, the loading pattern for the general factor indicated that it was indeed not general at all, as only the NA items exhibited substantial positive loadings ($\lambda \geq .636$). All loadings on the PA items were negative and negligible ($|\lambda| \leq .179$) except for item 3 (i.e., *excited*) that exhibited a small but positive loading ($\lambda = .140$). The loading pattern of the bi-factor model closely resembles a correlated two-factor model including a cross-loading for item 3. Since the bi-factor solution was not interpretable, we concluded, that Model 1c is the best representation of the PANAS structure for non-English samples. Factor loadings for Model 1c are presented in Figure 3 for both subgroups respectively.

Measurement Invariance Analysis

To analyze whether factor loadings are equal across groups, we compared the fit indices of freely estimated models to models with equality constraints on factor loadings of the PA and the NA items. The χ^2- difference test showed that the factor loadings cannot be considered equal between English and non-English samples for the PA ($\Delta\chi^2_{10} = 685.06$; $p < .001$) nor the NA factor ($\Delta\chi^2_{10} = 403.79$; $p < .001$). However, given conventional standards regarding differences in CFI (Meade, Johnson, & Braddy, 2008) only the differences in the PA factor seem to be practically important (PA: ΔCFI = .003; NA: ΔCFI = .002). On average, factor loadings are smaller in non-English samples (mean $\Delta\beta = .026$), which is especially true for PA items (mean $\Delta\beta = .055$). However, most deviations were quite small, except for item 3 (i.e., *excited*; $\Delta\beta = .176$), item 10 (i.e., *proud*; $\Delta\beta = .095$), and item 14 (i.e., *inspired*; $\Delta\beta = .090$).

Sensitivity Analyses

An anonymous reviewer correctly noticed that, of course, included studies differ in characteristics beyond the language of the PANAS that might also have an impact on the measurement model. Specifically, he or she mentioned that studies found that scale correlations change as a function of the time frame included in the instruction with decreased intercorrelations for measures of trait affect (Diener & Emmons, 1984; Krohne et al., 1996; Schmuckle, Egloff, & Burns, 2002). Thus, to test the robustness of the identified factor structure we repeated the MASEM analyses but excluded samples that used a trait instruction (i.e., asked participants how they feel in general). The corresponding correlation matrices closely resembled the previously derived correlation matrices with only minor differences in correlations for both the English ($M(|\Delta r|$ = .014, SD = .009 , Max = .041) and the non-English samples ($M(|\Delta r|$ = .002, SD = .001, Max = .006). Accordingly, differences in the model fit evaluations as well as differences in factor loadings for Model 1c were negligible (see supplementary material).

Discussion

In response to the ongoing discussion regarding its internal structure, the general purpose of our study was to provide a meta-analytic investigation of the PANAS dimensionality. Additionally, the robustness of factor solutions across different application contexts was tested. Results indicated a suboptimal model fit for the orthogonal two-factor model originally proposed by Watson and Tellegen (1985), whereas a correlated two-factor model including error correlations within content categories provided the best fit for both English and non-English samples. The practice of allowing measurement errors to correlate is controversially discussed within the structural equation modeling literature and many authors have argued against it (e.g., Cortina, 2002; Gerbing & Anderson, 1984; Hermida, 2015; Tomarken & Waller, 2003). Especially when the decision to include correlated uniquenesses is based on *post hoc* modifications it is considered inappropriate they might mask an actual underlying structure. However, for the PANAS, correlated errors are deduced *a priori* and are theoretically justified by the content categories defined by Zevon and Tellegen (1982). Of course, the correlated errors still point to substantial overlap in item content and a more complex measurement model underlying the responses (i.e., a hierarchical model). If one is interested in an affect measure that more closely resembles the proposed hierarchical taxonomic scheme, the expanded form of the PANAS (PANAS-X, Watson & Clark, 1994) would be a good choice. The PANAS-X is an extension of the original PANAS and contains 60 items that are assigned to the two higher-order factors NA and PA, which are respectively composed of distinguishable mood categories.

Although configural invariance was established, a closer investigation of the factor loadings and the results of the metric invariance tests indicated that the factor structure of the original PANAS items is not comparable to translated versions. Especially the PA items showed substantial differences and exhibited in general smaller loadings in non-English samples compared to English-speaking samples. Given that our moderator variable only differentiated English and translated versions, there might be two different explanations for this. First, differences in factor loadings between English and non-English samples might trace to differences in the general procedure and/or the quality of the translation process. This seems reasonable because many PANAS adaptations represent ad hoc translations. Only some studies explicitly follow common standards regarding the translation process and the proof of validity (e.g., Engelen et al., 2006; Krohne et al., 1996). An alternative explanation concerns cultural differences in the interpretation of the PANAS items. In this vein, some items of the PANAS might reflect an American or Western understanding of positive affect that might not reflect the positive emotional experience in other cultures. Indeed, some of the present results strongly point to this explanation. First, it seems reasonable that differences are more pronounced for PA items than for the NA scale, as cross-cultural comparisons show variations in the interpretation of positive emotions but not negative ones (Leu, Wang, & Koo, 2011; Uchida & Kitayama, 2009). This might be especially true for emotions that are associated with a high level of arousal (like excitement) as those are valued more in western countries than in eastern cultures (see Lim, 2016). This cultural explanation of the differences is further supported by the fact that especially those items are problematic that have been previously shown to be culturally ambiguous, like item 3 (i.e., excited) and item 10 (i.e., proud). The former seems to be the most ambiguous item for non-English samples, as it positively correlates with both PA and NA items (see Table 2), whereas the latter exhibits decreased factor loadings for non-English samples. In accordance, previous studies have shown that pride is barely associated with PA in eastern cultures (Diener, Oishi, & Tay, 2018; Kitayama, Markus, & Kurokawa, 2000; Scollon, Diener, Oishi, & Biswas-Diener, 2009). Additionally, our results indicate that the intercorrelation between PA and NA was smaller for translated versions than for the original English version of the PANAS. This can be an additional sign of cultural differences in response styles that have also been found in previous studies (Bagozzi, Wong, & Yi, 2010; Schimmack, Oishi, & Diener, 2002). It has been hypothesized that Eastern cultures promote a dialectic way of thinking that considers positive and negative affect as compatible, whereas Western cultures see them in oppositional ways. In this vein, correlations between PA and NA are expected to be decreased in Eastern compared to Western cultures.

Limitations

Even though this study is the first meta-analytic investigation of the PANAS structure, it still has several limitations. Since our meta-analytic approach is based on inter-item correlations that are solemnly presented in scientific articles, our final sample includes only a limited number of different countries. Especially, eastern countries are underrepresented. Since we argue that differences might reflect cultural differences in responses, the result of the moderator analysis presented here might underestimate actual differences between original and translated PANAS versions. Additionally, one might suggest that there might be further potential moderator variables besides the language of the PANAS. For instance, previous studies have evaluated the PANAS responses as a function of participants' age and found greater correlations between PA and NA among elderly participants (Kercher, 1992).

Thus, the analysis of further variables seems important to test the robustness of our results. The present sensitivity analysis presents only a first step and needs further support by means of additional subgroup or moderator analyses. However, the present data set is so far too small to accomplish this. Nevertheless, since all data is openly accessible, the data set can be extended by future samples to analyze the influence of further categorical variables. Most recently, an adaptation of the analysis approach to evaluate the influence of continuous moderators has also been developed (Jak & Cheung, 2020).

Conclusion

In conclusion, our results illustrate that although PA and NA form distinct but interrelated factors in different cultures and languages, the adjectives that are associated with PA are used differently. Although it is not possible to distinguish whether these differences trace back to differences in the quality of the translation process or whether they are attributable to cultural differences between samples, the implication is the same: Cross-national research with the PANAS might be justifiable only if measurement equivalence has been sufficiently demonstrated for the nations at hand. Accordingly, more studies are needed that explicitly analyze measurement invariance of the PANAS.

References

Note: References marked with an asterisk are included in the meta-analysis

Akhter, N. (2017). Urdu translation and validation of shorter version of Positive Affect and Negative Affect Schedule (PANAS) on Pakistani bank employees. *JPMA. The Journal of the Pakistan Medical Association*, *67*(10), 1517–1523.

*Anvari, F., & Lakens, D. (2019). *Using Anchor-Based Methods to Determine the Smallest Effect Size of Interest.* https://doi.org/10.31234/osf.io/syp5a

*Baltes, M., & Lang, F. (2012). *Altern und Lebenserfahrung im Erwachsenenalter (ALLEE): Forschungsdaten [Data file]*. https://doi.org/10.5160/PSYCHDATA.LGFR97AL01

*Barthels, F., Barrada, J. R., & Roncero, M. (2019). Orthorexia nervosa and healthy orthorexia as new eating styles. *PLoS ONE*, *14*(7), e0219609. https://doi.org/10.1371/journal.pone.0219609

*Bastian, B., Jetten, J., & Ferris, L. J. (2014). Pain as social glue: Shared pain increases cooperation. *Psychological Science*, *25*(11), 2079–2085. https://doi.org/10.1177/0956797614545886

Beck, J. G., Novy, D. M., Diefenbach, G. J., Stanley, M. A., Averill, P. M., & Swann, A. C. (2003). Differentiating anxiety and depression in older adults with generalized anxiety disorder. *Psychological Assessment*, *15*(2), 184–192. https://doi.org/10.1037/1040-3590.15.2.184

*Bilali, R., Iqbal, Y., & Erisen, C. (2019). The role of lay beliefs about group transgressions in acceptance of responsibility for ingroup harm-doing. *European Journal of Social Psychology*, *49*(5), 992–1006. https://doi.org/10.1002/ejsp.2583

*Caicedo Cavagnis, E., Michelini, Y., Belaus, A., Mola, D. J., Godoy, J. C., & Reyna, C. (2018). Further considerations regarding PANAS: Contributions from four studies with different Argentinean samples. *Suma Psicológica*, *25*(2), 133–145. https://doi.org/10.14349/sumapsi.2018.v25.n2.5

*CentERdata (2020a). *LISS Immigrant Sample Data: Wave 1 - Wave 2 [Data file]*. Retrieved from www.lissdata.nl

*CentERdata (2020b). *LISS Panel Data: Wave 1 - Wave 9 [Data file]*. Retrieved from www.lissdata.nl

Cheung, M. W. L. (2014). Metasem: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology*, *5*, 1521. https://doi.org/10.3389/fpsyg.2014.01521

Cheung, M. W. L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods*, *10*(1), 40–64. https://doi.org/10.1037/1082-989X.10.1.40

Cortina, J. M. (2002). Big things have small beginnings: An assortment of “minor” methodological misunderstandings. *Journal of Management*, *28*(3), 339-362. https://doi.org/10.1177/014906302028003055

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. *The 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., . . . Botella, C. (2018). *Positive And Negative Affect Schedule (PANAS): Psychometric Properties For The Online Version In A Clinical Sample With Emotional Disorders [Data file]*. Retrieved from https://doi.org/10.5281/ZENODO.1477084

*Dickinson, D. L., Drummond, S. P. A., & McElroy, T. (2017). The viability of an ecologically valid chronic sleep restriction and circadian timing protocol: An examination of sample attrition, compliance, and effectiveness at impacting sleepiness and mood. *PLoS ONE*, *12*(3), e0174367. https://doi.org/10.1371/journal.pone.0174367

Diener, E., & Emmons, R. A. (1984). The independence of positive and negative affect. *Journal of Personality and Social Psychology*, *47*, 1105-1117. https://doi.org/10.1037/0022-3514.47.5.1105

Diener, E., Oishi, S., & Tay, L. (Eds.) (2018). *Handbook of Well-Being*. Salt Lake City: DEF Publishers.

Egloff, B., Schmukle, S. C., Burns, L. R., Kohlmann, C. W., & Hock, M. (2003). Facets of dynamic positive affect: Differentiating joy, interest, and activation in the positive and negative affect schedule (PANAS). *Journal of Personality and Social Psychology*, *85*(3), 528–540. https://doi.org/10.1037/0022-3514.85.3.528

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods*, *22*(3), 541–562. https://doi.org/10.1037/met0000083

Engelen, U., Peuter, S. de, van Diest, I., & van den Bergh, O. (2006). Further validation of the Positive and Negative Affect Schedule (PANAS) and comparision of two Dutch versions. *Gedrag & Gezondheid*, *34*, 61–70. https://doi.org/10.1007/BF03087979

*Estévez-López, F., Pulido-Martos, M., Armitage, C. J., Wearden, A., Álvarez-Gallardo, I. C., Arrayás-Grajera, M. J., . . . 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

Galinha, I. C., Pereira, C. R., & Esteves, F. G. (2013). Confirmatory factor analysis and temporal invariance of the Positive and Negative Affect Schedule (PANAS). *Psicologia: Reflexão E Crítica*, *26*(4), 671–679. https://doi.org/10.1590/S0102-79722013000400007

Gaudreau, P., Sanchez, X., & Blondin, J. P. (2006). Positive and Negative Affective States in a Performance-Related Setting. *European Journal of Psychological Assessment*, *22*(4), 240–249. https://doi.org/10.1027/1015-5759.22.4.240

Gençöz, T. (2000). Positive and Negative Affect Schedule: A study of validity and reliability. *Türk Psikoloji Dergisi*, *15*(46), 19–28.

*GESIS (2020). *GESIS Panel - Standard Edition.* https://doi.org/10.4232/1.13436

Gerbing, D. W., & Anderson, J. C. (1984). On the meaning of within-factor correlated measurement errors. *Journal of Consumer Research, 11*(1), 572-580. https://doi.org/10.1086/208993

*Giakoumaki, S. G., Karagiannopoulou, L., Rózsa, S., Zouraraki, C., Karamaouna, P., & Cloninger, C. R. (2016). Psychometric properties of the Greek TCI-R and its clinical correlates: Schizotypy and the self-regulation of affective and cognitive functioning. *PeerJ, 4*, e1830. https://doi.org/10.7717/peerj.1830

*Greenland, K. (2010). *The role of physiological arousal and intergroup anxiety in prejudice [Data file]*. Colchester, Essex. https://doi.org/10.5255/UKDA-SN-850413

*Hameister, N., Kausmann, C., Müller, D., Simonson, J., Vogel, C., & Tesch-Römer, C. (2019). *Scientific Use File Deutscher Freiwilligensurvey (SUF FWS) 2014, Version 1.3 [Data file]*. https://doi.org/10.5156/FWS.2014.M.004

*Hameister, N., Müller, D., & Kausmann, C. (2019a). *Scientific Use File Deutscher Freiwilligensurvey (SUF FWS) 1999, Version 3.2 [Data File].* https://doi.org/10.5156/FWS.1999.M.003

*Hameister, N., Müller, D., & Kausmann, C. (2019b). *Scientific Use File Deutscher Freiwilligensurvey (SUF FWS) 2009, Version 3.2 [Data file].* https://doi.org/10.5156/FWS.2009.M.003

Hermida, R. (2015). The problem of allowing correlated errors in structural equation modeling: Concerns and considerations. *Computational Methods in Social Sciences*, *3*(1), 5–17.

Heubeck, B. G., & Wilkinson, R. (2019). Is all fit that glitters gold? Comparisons of two, three and bi-factor models for Watson, Clark & Tellegen's 20-item state and trait PANAS. *Personality and Individual Differences*, *144*, 132–140. https://doi.org/10.1016/j.paid.2019.03.002

*Horwood, S., & Anglim, J. (2019). Problematic smartphone usage and subjective and psychological well-being. *Computers in Human Behavior*, *97*, 44–50. https://doi.org/10.1016/j.chb.2019.02.028

*Huynh, A. C., Oakes, H., Shay, G. R., & McGregor, I. (2017). The wisdom in virtue: Pursuit of virtue predicts wise reasoning about personal conflicts. *Psychological Science*, *28*(12), 1848–1856. https://doi.org/10.1177/0956797617722621

Jackson, T., & Chen, H. (2008). Predicting changes in eating disorder symptoms among Chinese adolescents: A 9-month prospective study. *Journal of Psychosomatic Research*, *64*(1), 87–95. https://doi.org/10.1016/j.jpsychores.2007.08.015

Jak, S. (2015). *Meta-Analytic Structural Equation Modelling* (1st ed. 2015). Springer. https://doi.org/10.1007/978-3-319-27174-3

Jak, S., & Cheung, M. W. L. (2018a). Testing moderator hypotheses in meta-analytic structural equation modeling using subgroup analysis. *Behavior Research Methods*, *50*(4), 1359–1373. https://doi.org/10.3758/s13428-018-1046-3

Jak, S., & Cheung, M. W. L. (2018b). Accounting for missing correlation coefficients in fixed-effects MASEM. *Multivariate Behavioral Research*, *53*(1), 1-4. https://doi.org/10.1080/00273171.2017.1375886

Jak, S., & Cheung, M. W. L. (2020). Meta-analytic structural equation modeling with moderating effects on SEM parameters. *Psychological Methods, 25*(4), 430-455. https://doi.org/10.1037/met0000245

Joiner, T. E., Sandín, 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

Kemmelmeier, M. (2016). Cultural differences in survey responding: Issues and insights in the study of response biases. *International Journal of Psychology : Journal International De Psychologie*, *51*(6), 439–444. https://doi.org/10.1002/ijop.12386

Kercher, K. (1992). Assessing Subjective Well-Being in the Old-Old. *Research on Aging*, *14*(2), 131–168. https://doi.org/10.1177/0164027592142001

Killgore, W. D. S. (2000). Evidence for a third factor on the positive and negative affect schedule in a college student sample. *Perceptual and Motor Skills*, *90*(1), 147. https://doi.org/10.2466/PMS.90.1.147-152

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

Krohne, H. W., Egloff, B., Kohlmann, C. W., & Tausch, A. (1996). Investigations with a German version of the positive and negative affect schedule. *Diagnostica*, *42*(2), 139–156.

*Laboratory for the Study of Stress, Immunity, and Disease (2016a). *Common Cold Project: British Cold Study [Data file]*. Retrieved from www.commoncoldproject.com

*Laboratory for the Study of Stress, Immunity, and Disease (2016b). *Common Cold Project: Pittsburgh Cold Study 3 [Data file]*. Retrieved from http://www.commoncoldproject.com

Laurent, J., Catanzaro, S. J., Joiner, T. E., Rudolph, K. D., Potter, K. I., Lambert, S., . . . Gathright, T. (1999). A measure of positive and negative affect for children: Scale development and preliminary validation. *Psychological Assessment*, *11*(3), 326–338. https://doi.org/10.1037/1040-3590.11.3.326

*Lee, E. M., Klement, K. R., Ambler, J. K., Loewald, T., Comber, E. M., Hanson, S. A., . . . Sagarin, B. J. (2016). Altered States of Consciousness during an Extreme Ritual. *PLoS ONE*, *11*(5), e0153126. https://doi.org/10.1371/journal.pone.0153126

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

Leu, J., Wang, J., & Koo, K. (2011). Are positive emotions just as “positive” across cultures? *Emotion (Washington, D.C.)*, *11*(4), 994–999. https://doi.org/10.1037/a0021332

Leue, A., & Beauducel, A. (2011). The PANAS structure revisited: On the validity of a bifactor model in community and forensic samples. *Psychological Assessment*, *23*(1), 215–225. https://doi.org/10.1037/a0021400

Leue, A., & Lange, S. (2011). Reliability generalization: An examination of the Positive Affect and Negative Affect Schedule. *Assessment*, *18*(4), 487–501. https://doi.org/10.1177/1073191110374917

Lim, N. (2016). Cultural differences in emotion: Differences in emotional arousal level between the East and the West. *Integrative Medicine Research*, *5*(2), 105-109. https://doi.org/10.1016/j.imr.2016.03.004

Mackinnon, A., Jorm, A. F., Christensen, H., Korten, A. E., Jacomb, P. A., & Rodgers, B. (1999). A short form of the Positive and Negative Affect Schedule: evaluation of factorial validity and invariance across demographic variables in a community sample. *Personality* *and Individual Differences*, *27*(3), 405–416. https://doi.org/10.1016/S0191-8869(98)00251-7

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

*Medvedev, O. N., & Landhuis, C. E. (2018). Exploring constructs of well-being, happiness and quality of life. *PeerJ*, *6*, e4903. https://doi.org/10.7717/peerj.4903

Mehrabian, A. (1997). Comparison of the PAD and PANAS as models for describing emotions and for differentiating anxiety from depression. *Journal of Psychopathology and Behavioral Assessment*, *19*, 331–357. https://doi.org/10.1007/BF02229025

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

Merz, E. L., & Roesch, S. C. (2011). Modeling trait and state variation using multilevel factor analysis with PANAS daily diary data. *Journal of Research in Personality*, *45*(1), 2–9. https://doi.org/10.1016/j.jrp.2010.11.003

Mihic, L., Novovic, Z., Colovic, P., & Smederevac, S. (2014). Serbian adaptation of the positive and negative affect schedule (PANAS): Its facets and second-order structure. *Psihologija*, *47*(4), 393–414. https://doi.org/10.2298/PSI1404393M

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

*Müller, D., Hameister, N., & Kausmann, C. (2019). *Scientific Use File Deutscher Freiwilligensurvey (SUF FWS) 2004, Version 3.2 [Data file]*. https://doi.org/10.5156/FWS.2004.M.003

*Naragon-Gainey, K., & DeMarree, K. G. (2017). Decentering attenuates the associations of negative affect and positive affect with psychopathology. *Clinical Psychological Science*, *5*(6), 1027–1047. https://doi.org/10.1177/2167702617719758

Nye, C. D., Roberts, B. W., Saucier, G., & Zhou, X. (2008). Testing the measurement equivalence of personality adjective items across cultures. *Journal of Research in Personality*, *42*(6), 1524–1536. https://doi.org/10.1016/j.jrp.2008.07.004

*Ocampo Rebollar, A. (2017). *Positive and negative affect fluctuations during the menstrual cycle [Data file]*. https://doi.org/10.5255/UKDA-SN-852672

*Rahm, T., Heise, E., & Schuldt, M. (2017). Measuring the frequency of emotions-validation of the Scale of Positive and Negative Experience (SPANE) in Germany. *PLoS ONE*, *12*(2), e0171288. https://doi.org/10.1371/journal.pone.0171288

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. https://doi.org/10.1037/met0000045

Rush, J., & Hofer, S. M. (2014). Differences in within- and between-person factor structure of positive and negative affect: Analysis of two intensive measurement studies using multilevel structural equation modeling. *Psychological Assessment*, *26*(2), 462–473. https://doi.org/10.1037/a0035666

*Ryff, C. D., & Davidson, R. J. (2010). *Midlife in the United States (MIDUS 2): Neuroscience Project, 2004-2009 [Data file]*. https://doi.org/10.3886/ICPSR28683.V4

*Sailer, U., Rosenberg, P., Nima, A. A., Gamble, A., Gärling, T., Archer, T., & Garcia, D. (2014). A happier and less sinister past, a more hedonistic and less fatalistic present and a more structured future: Time perspective and well-being. *PeerJ*, *2*, e303. https://doi.org/10.7717/peerj.303

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

*Schindler, S., Pfattheicher, S., & Reinhard, M.-A. (2019). Potential negative consequences of mindfulness in the moral domain. *European Journal of Social Psychology*, *49*(5), 1055–1069. https://doi.org/10.1002/ejsp.2570

Schmuckle, S. C., Egloff, B., & Burns, L. R. (2002). The relationship between positive and negative affect in the Positive and Negative Affect Schedule. *Journal of Research in Personality*, *36*, 463-475. https://doi.org/10.1016/S0092-6566(02)00007-7

Scollon, C. N., Diener, E., Oishi, S., & Biswas-Diener, R. (2009). Emotions Across Cultures and Methods. In E. Diener (Ed.), *Social Indicators Research Series: Vol. 38. Culture and well-being* (Vol. 38, pp. 203–228). Springer. https://doi.org/10.1007/978-90-481-2352-0_10

Seib-Pfeifer, L. E., Pugnaghi, G., Beauducel, A., & Leue, A. (2017). On the replication of factor structures of the Positive and Negative Affect Schedule (PANAS). *Personality and Individual Differences*, *107*, 201–207. https://doi.org/10.1016/j.paid.2016.11.053

Sireci, S. G., Yang, Y., Harter, J., & Ehrlich, E. J. (2006). Evaluating Guidelines For Test Adaptations. *Journal of Cross-Cultural Psychology*, *37*(5), 557–567. https://doi.org/10.1177/0022022106290478

*Sutton-Tyrrell, K., Selzer, F., Sowers, M., Finkelstein, J., Powell, L., Gold, E., . . . Brooks, M. M. (2014). *Study of Women's Health Across the Nation (SWAN), 2006-2008: Visit 10 Dataset [Data file]*. https://doi.org/10.3886/ICPSR32961.V2

Terracciano, A., McCrae, R. R., & Costa, P. T., JR. (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

*Toet, A., Houtkamp, J. M., & Vreugdenhil, P. E. (2016). Effects of personal relevance and simulated darkness on the affective appraisal of a virtual environment. *PeerJ*, *4*, e1743. https://doi.org/10.7717/peerj.1743

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with „well fitting“ models. *Journal of Abnormal Psychology*, *112*(4), 578-598. https://doi.org/10.1037/0021-843X.112.4.578

*Traxler, J., Madden, V. J., Moseley, G. L., & Vlaeyen, J. W. S. (2019). Modulating pain thresholds through classical conditioning. *PeerJ*, *7*, e6486. https://doi.org/10.7717/peerj.6486

*Troup, L. J., Andrzejewski, J. A., Braunwalder, J. T., & Torrence, R. D. (2016). The relationship between cannabis use and measures of anxiety and depression in a sample of college campus cannabis users and non-users post state legalization in Colorado. *PeerJ*, *4*, e2782. https://doi.org/10.7717/peerj.2782

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

Uchida, Y., & Kitayama, S. (2009). Happiness and Unhappiness in East and West: Themes and variations. *Emotion*, *9*(4), 441-456. https://doi.org/10.1037/a0015634

*Van der Meij, L., Klauke, F., Moore, H. L., Ludwig, Y. S., Almela, M., & van Lange, P. A. M. (2015). Football fan aggression: The importance of low basal cortisol and a fair referee. *PLoS ONE*, *10*(4), e0120103. https://doi.org/10.1371/journal.pone.0120103

Vera-Villarroel, P., Urzúa, A., Jaime, D., Contreras, D., Zych, I., Celis-Atenas, K., . . . Lillo, S. (2019). Positive and Negative Affect Schedule (PANAS): Psychometric Properties and Discriminative Capacity in Several Chilean Samples. *Evaluation & the Health Professions*, *42*(4), 473–497. https://doi.org/10.1177/0163278717745344

Villodas, F., Villodas, M. T., & Roesch, S. (2011). Examining the Factor Structure of the Positive and Negative Affect Schedule (PANAS) in a Multiethnic Sample of Adolescents. *Measurement and Evaluation in Counseling and Development*, *44*(4), 193–203. https://doi.org/10.1177/0748175611414721

Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the Positive and Negative Affect Schedule*. Ames: University of Iowa. https://doi.org/10.13072/midss.438

Watson, D., & Clark, L. A. (1997). Measurement and mismeasurement of mood: Recurrent and emergent issues. *Journal of Personality Assessment*, *68*(2), 267–296. https://doi.org/10.1207/s15327752jpa6802_4

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

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*(2), 219–235. https://doi.org/10.1037/0033-2909.98.2.219

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology*, *76*(5), 820–838. https://doi.org/10.1037/0022-3514.76.5.820

*Wedderhoff, O., Chasiotis, A., & Rosman, T. (2019). *Don't Tell Me What to Think: How Perceived and Suggested Risk Affect Selective Exposure to Health Information [Data file]*. Retrieved from Leibniz Institute for Psychology Information (ZPID) website: https://www.psycharchives.org/. https://doi.org/10.23668/PSYCHARCHIVES.2435

*Zampetakis, L. A., Lerakis, M., Kafetsios, K., & Moustakis, V. (2015). Using item response theory to investigate the structure of anticipated affect: Do self-reports about future affective reactions conform to typical or maximal models? *Frontiers in Psychology*, *6*, 1438. https://doi.org/10.3389/fpsyg.2015.01438

Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology*, *43*(1), 111–122. https://doi.org/10.1037/0022-3514.43.1.111

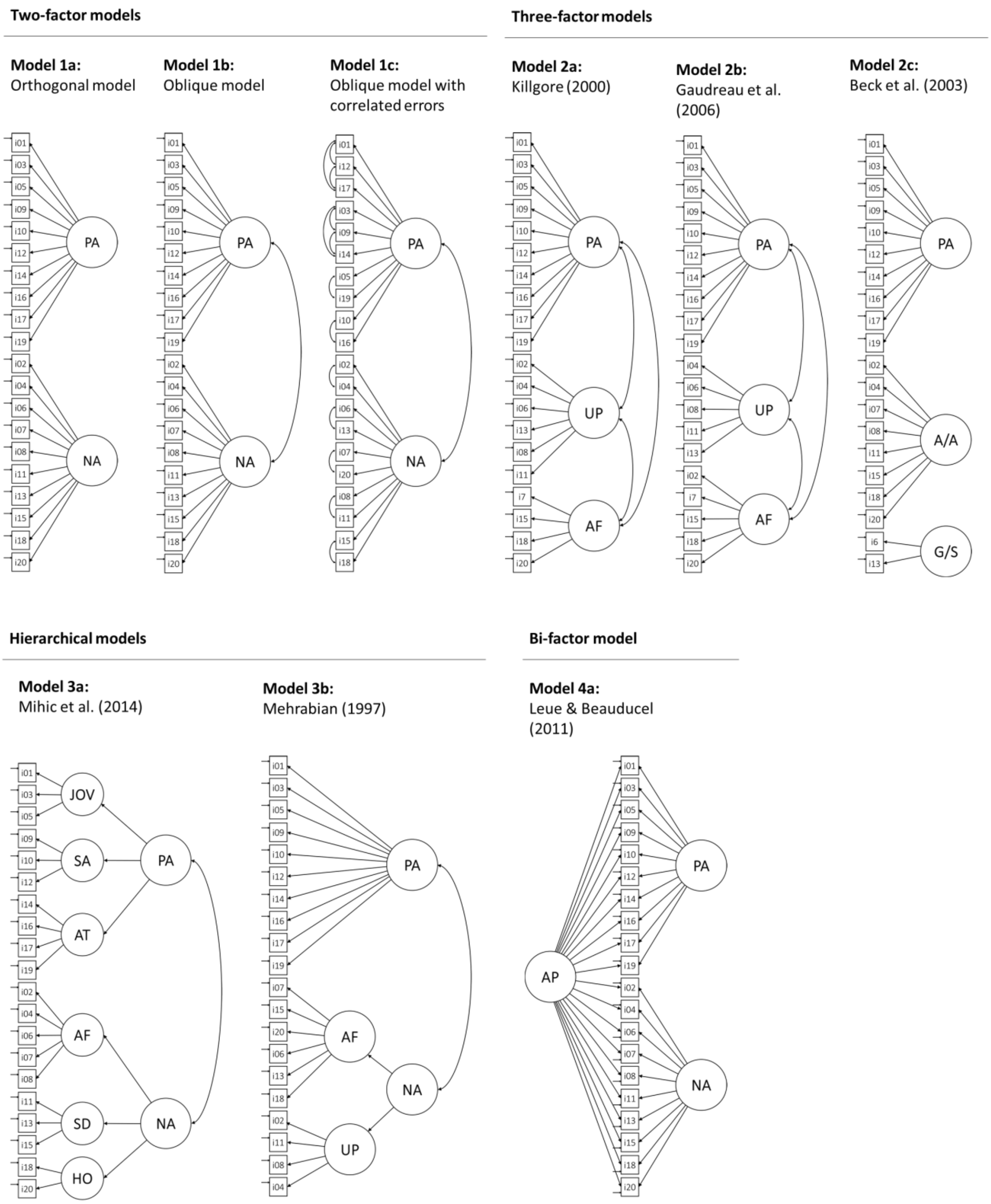

Figure 1. Graphical representation of competing measurement models. PA = positive affect, NA = negative affect, AF = afraid, UP = upset, JOV = joviality, SA = self-assurance, AT = attentiveness, SD = self-disgust, HO = hostility, A/A = anxiety and anger, G/S = guilt and shame, AP = affective polarity

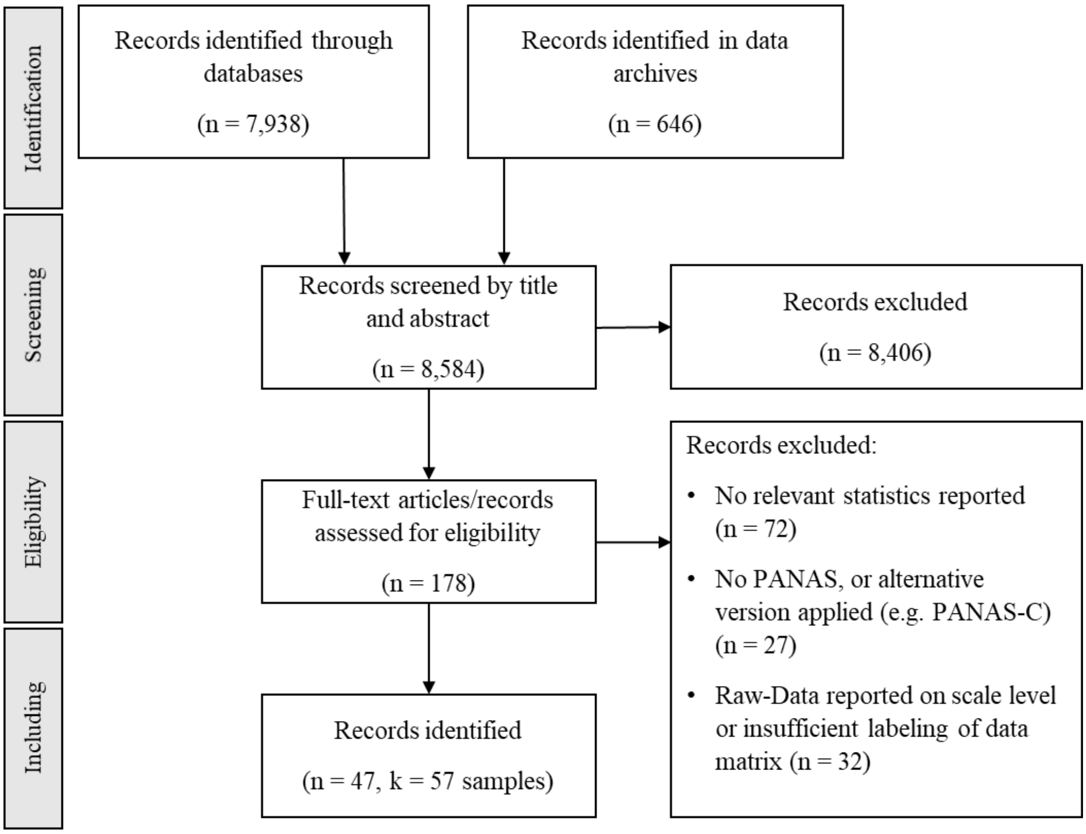

Figure 2: PRISMA flow chart of the literature search and the selection process

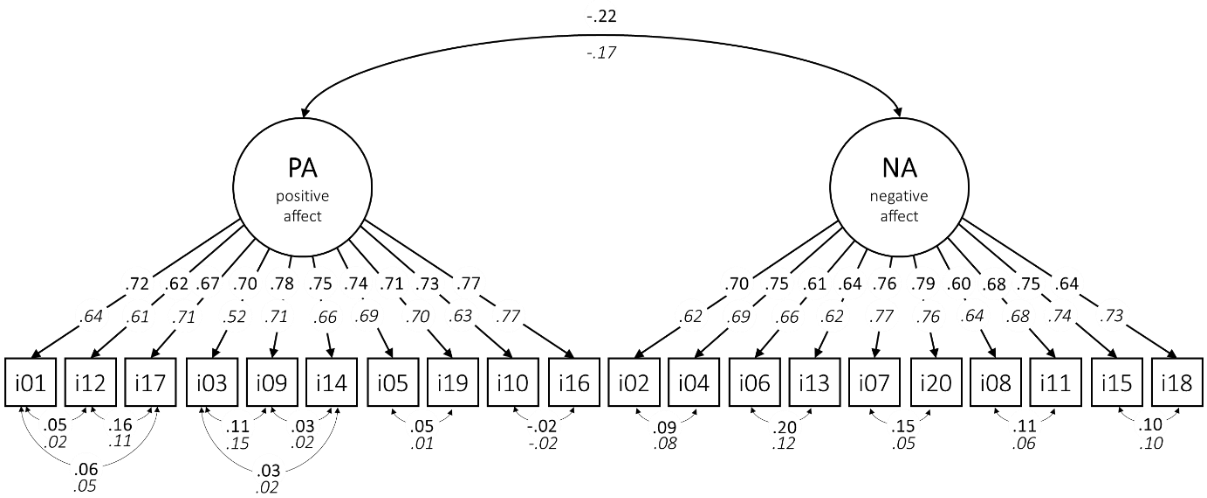

Figure 3. Factor loadings for the original PANAS items (**bold**) and translated PANAS items (*italics*)

Table 1

PANAS items with assignment to affective dimensions and content categories

Nr.	PANAS items	Dimension	Content category[a]
i01	interested	PA	attentive
i02	distressed	NA	distressed
i03	excited	PA	excited
i04	upset	NA	distressed
i05	strong	PA	strong
i06	guilty	NA	guilty
i07	scared	NA	fearful
i08	hostile	NA	angry
i09	enthusiastic	PA	excited
i10	proud	PA	proud
i11	irritable	NA	angry
i12	alert	PA	attentive
i13	ashamed	NA	guilty
i14	inspired	PA	excited
i15	nervous	NA	jittery
i16	determined	PA	proud
i17	attentive	PA	attentive
i18	jittery	NA	jittery
i19	active	PA	strong
i20	afraid	NA	fearful

Note. PA = positive affect; NA = negative affect

[a]proposed by Zevon & Tellegen (1982)

Table 2

Pooled correlation matrix for the original PANAS items (lower triangle) and translated PANAS items (upper triangle)

	i01	i02	i03	i04	i05	i06	i07	i08	i09	i10	i11	i12	i13	i14	i15	i16	i17	i18	i19	i20
i01	1	-.063	.306	-.042	.369	-.058	-.098	-.093	.423	.323	-.052	.375	-.065	.409	-.043	.405	.446	-.109	.420	-.113
i02	-.113	1	.021	.460	-.152	.355	.404	.303	-.134	-.040	.389	-.073	.322	-.070	.442	-.126	-.090	.441	-.126	.450
i03	.514	-.066	1	.124	.346	.109	.095	.090	.492	.327	.082	.237	.124	.349	.093	.303	.242	.079	.308	.042
i04	-.126	.593	-.045	1	-.055	.404	.474	.438	-.057	.010	.510	-.016	.366	.004	.464	-.042	-.046	.439	-.052	.417
i05	.431	-.155	.442	-.139	1	-.060	-.121	-.045	.465	.442	-.072	.365	-.069	.376	-.122	.458	.368	-.146	.447	-.177
i06	-.066	.381	-.002	.421	-.080	1	.540	.435	-.018	-.036	.368	-.051	.520	.031	.374	-.083	-.087	.401	-.066	.420
i07	-.085	.467	.004	.485	-.119	.459	1	.506	-.059	-.042	.390	-.063	.487	-.004	.479	-.124	-.105	.497	-.108	.579
i08	-.103	.388	-.006	.447	-.057	.359	.384	1	-.037	.016	.435	-.049	.410	.011	.367	-.049	-.090	.374	-.072	.382
i09	.557	-.156	.634	-.149	.488	-.073	-.081	-.058	1	.475	-.049	.345	.005	.454	-.043	.440	.376	-.082	.439	-.093
i10	.402	-.156	.472	-.133	.530	-.068	-.077	-.033	.526	1	.042	.292	-.004	.343	.000	.385	.318	-.029	.343	-.058
i11	-.162	.462	-.077	.535	-.133	.346	.363	.492	-.148	-.104	1	-.011	.377	-.008	.512	-.042	-.054	.449	-.069	.401
i12	.410	-.101	.316	-.097	.426	-.070	-.060	-.049	.390	.371	-.084	1	-.030	.392	-.025	.447	.528	-.095	.431	-.076
i13	-.086	.401	-.025	.425	-.126	.575	.468	.362	-.084	-.094	.352	-.061	1	.061	.405	-.067	-.067	.420	-.063	.443
i14	.484	-.104	.515	-.092	.475	-.005	-.022	-.042	.567	.512	-.129	.385	-.030	1	.038	.452	.406	-.029	.439	-.032
i15	-.058	.482	.026	.462	-.142	.386	.581	.330	-.078	-.086	.418	-.072	.415	-.029	1	-.069	-.046	.620	-.062	.552
i16	.461	-.095	.419	-.077	.499	-.084	-.050	-.070	.495	.481	-.076	.464	-.109	.554	-.024	1	.574	-.128	.492	-.152
i17	.477	-.127	.350	-.134	.425	-.096	-.082	-.099	.456	.396	-.153	.550	-.108	.422	-.079	.517	1	-.119	.480	-.102
i18	-.063	.411	.033	.379	-.102	.339	.448	.356	-.037	-.051	.401	-.041	.376	-.021	.556	-.026	-.058	1	-.120	.576
i19	.433	-.115	.448	-.120	.520	-.065	-.078	-.050	.508	.443	-.125	.444	-.085	.457	-.072	.486	.461	-.009	1	-.125
i20	-.109	.475	-.028	.485	-.154	.437	.743	.371	-.109	-.104	.384	-.071	.488	-.048	.607	-.068	-.104	.481	-.074	1

Table 3

Fit indices and information criteria for nine competing CFA models for samples using the original or the translated version of the PANAS

		original version					translated version				
		CFI	*SRMR*	*RMSEA*	*AIC*	*BIC*	*CFI*	*SRMR*	*RMSEA*	*AIC*	*BIC*
Two-factor models											
1a	Orthogonal two-factor	.898	.104	.065	5,186.707	4,010.150	.864	.084	.061	29,392.315	27,906.928
1b	Oblique two-factor	.907	.100	.062	4,751.284	3,581.648	.868	.079	.061	28,434.016	26,957.367
1c	with correlated errors	.952	.059	.047	2,399.112	1,319.448	.904	.065	.054	20,610.256	19,247.190
Three-factor models											
2a	Killgore (2000)	.925	.082	.056	3,797.380	2,641.586	.878	.075	.059	26,328.411	24,869.237
2b	Beck et al. (2003)	.765	.154	.099	12,202.505	11,025.948	.657	.159	.097	74,058.745	72,573.359
2c	Gaudreau et al. (2006)	.917	.089	.059	4,222.126	3,066.332	.879	.075	.058	26,079.052	24,619.878
Second-order models											
3a	Mihic et al. (2013)	.909	.099	.063	4,655.310	3,527.199	.877	.075	.060	26,561.145	25,136.921
3b	Mehrabian (1997)	.923	.083	.057	3,911.833	2,756.039	.876	.076	.059	26,685.010	25,225.836
Bi-factor model											
4a	Leue & Beauducel (2011)	.935	.076	.055	3,302.981	2,264.843	.916	.060	.051	18,038.909	16,728.274