Longitudinal and sex measurement invariance of the Affective Neuroscience Personality Scales

Abstract

The Affective Neuroscience Personality Scales (ANPS) is a personality instrument based on six evolutionary-related brain systems that are at the foundation of human emotions and behaviors: SEEKING, CARING, PLAYFULNESS, FEAR, ANGER, and SADNESS. We sought to assess for the short and long versions of the ANPS: (i) the longitudinal measurement invariance and long-term (4-year) stability, and (ii) the sex measurement invariance. Using data from a Canadian cohort (N=518), we used single-group confirmatory factor analysis (CFA) to assess longitudinal invariance and multiple-group CFA to assess sex invariance, according to a five-step approach evaluating five invariance levels (configural, metric, scalar, factorial, and complete). Results supported full longitudinal invariance for both versions for all invariance levels. Partial residual invariance was supported for sex invariance. The long-term stability of both versions was good to excellent. Implications for personality assessment and ANPS development are discussed.

Keywords

Emotions, ANPS, SEEKING, CARING, PLAYFULNESS, FEAR, ANGER, SADNESS, personality assessment, measurement invariance,

Introduction

The Affective Neuroscience Personality Scales (ANPS) (Davis, Panksepp, & Normansell, 2003) is a self-report questionnaire designed to assess emotional dispositions related to activity in primary-process affective networks and associated hormones. These primary affective networks mold the development of higher-order mental skills and frame the individual's subjective feelings, behaviors, and relationships (Panksepp, 2006; Davis & Panksepp, 2011; Panksepp, 2007; Panksepp & Panksepp, 2013). Each ANPS subscale is based on ethological research and neurobiological studies that point towards at least six evolutionary-related brain and behavioral core systems at the foundation of human emotions and behaviors (Panksepp, 1998, 2005, 2006; Panksepp & Biven, 2012; Toronchuk & Ellis, 2013).

These systems correspond to three positive and three negative emotional systems (Panksepp, 1998, 2005; upper-case letters refer to the systems in Panksepp's model and are followed by their behavioral counterparts in humans): (1) SEEKING/interest (being curious, exploring, striving for solutions to problems, positively anticipating new experiences), (2) PLAYFULNESS/joy (having fun, playing games with physical contact, humor, and laughter), (3) CARING/nurturance (being drawn to young children and pets, feeling softhearted toward animals and people in need, feeling empathy), (4) ANGER/rage (feeling hotheaded, being easily irritated and frustrated, experiencing frustration leading to anger, expressing anger verbally or physically), (5) FEAR/anxiety (feeling tense, worrying, struggling with decisions, ruminating), (6) SADNESS/panic and separation distress (feeling lonely, crying frequently, thinking about loved ones and past relationships, and feeling distress).

The ANPS were modeled in the belief that an accurate questionnaire for assessing emotional personality should aim to “carve personality along the lines of emerging brain systems that help generate the relevant psychological attributes” (Davis et al., 2003, p 58; see also: Cloninger, 1987; Cloninger, Svrakic, & Przybeck, 1993; Ersche, Turton, Pradhan, Bullmore, & Robbins, 2010; Gray, 1987). In this respect, the underpinnings of the ANPS differ from those of personality scales relying on the Five-Factor Model (FFM). The FFM is based on a lexical hypothesis positing that “most of the socially relevant and salient personality characteristics have become encoded in the natural language” (John & Srivastava, 2001; p. 103). According to this approach, the most relevant aspects that differentiate groups of people appear verbally (Saucier, 2009). The FFM nonetheless focuses on phenotypic characteristics of personality (John & Srivastava, 2001), and measures of personality that better reflect underlying biological processes are still needed (see also Montag & Reuter, 2014; Reuter, Cooper, Smillie, Markett, & Montag, 2015, for a recent discussion on the advantages of the ANPS over the FFM for investigating the molecular genetic bases of personality).

More than a decade ago, Gottesman & Gould (2003) defined endophenotypes as “measurable components unseen by the unaided eye along the pathway between disease and distal genotype,” p. 636); they suggested endophenotypes could further enhance our understanding of the underlying mechanisms of mental illnesses by reducing the gap between underlying biological processes and behavior. The endophenotypic approach is considered a solution for circumventing the limitations of the current diagnostic systems for mental disorders, which do not seem to have optimally assisted the search for disorder-specific pathophysiological mechanisms or biological and cognitive markers (McGorry & van Os, 2013). Several psychiatric disorders share common emotional deficits and associated cerebral patterns (Kret & Ploeger, 2015; Kelley, Wagner, & Heatherton, 2015). For instance, social phobia (Axis I) and avoidant personality disorder (Axis II) present similar characteristics and share both psychological and biological processes pertaining to emotional regulatory functions (Siever & Weinstein, 2009; Stein & Stein, 2008). The phenotypic heterogeneity of disorders and the overlap between different diagnostic entities are major limitations to the advance of knowledge in this field (McGorry & Nelson, 2016), and many researchers are now seeking other theoretical and heuristic models (Kendler, Zachar, & Craver, 2011; Krueger & Eaton, 2015) (see also the Research Domain Criteria [RDoC] project; Maj, 2014). The dimensional conceptualization of personality disorders in the latest edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) is an example of this ongoing paradigm shift (Krueger & Markon, 2014).

In this scientific context, the ANPS may be a useful transdiagnostic tool that could enable a more fine-grained evaluation of the emotional and motivational difficulties present in many psychiatric disorders, and an increasing number of studies now use this instrument. ANPS scores have been related to both genetic (e.g., FEAR and SADNESS with the serotonin transporter polymorphism and the oxytocin receptor gene markers; ANGER with the dopaminergic polymorphism) and neurobiological substrates (e.g., a negative association between ANGER or FEAR scores and amygdala volume; Berthoz, Orvoën, & Grezes, 2010; Felten, Montag, Markett, Walter, & Reuter, 2011; Montag & Reuter, 2014; Montag, Reuter, Jurkiewicz, Markett, & Panksepp, 2013; Reuter, Weber, Fiebach, Elger, & Montag, 2009). In addition to neurobiological studies, Pingault et al. offered evidence of the validity of the ANPS based on its relations with other variables. They reported, for example, positive associations between ANGER/rage and Multidimensional Anger Inventory scores, between FEAR/anxiety and Spielberger State Trait Anxiety Inventory trait scores, between SADNESS/panic and Beck Depression Inventory scores (Pingault, Pouga, Grèzes, & Berthoz, 2012). The ANPS is also being used in clinical settings, for example, among patients with neurological (Farinelli et al., 2013, 2015) and psychiatric disorders (Savitz, Van der Merwe, & Ramesar, 2008a; J. Savitz, Van Der Merwe, & Ramesar, 2008b). Only three studies have explored the convergent validity between the ANPS and FFM measures, in American (Davis et al., 2003), Turkish (Özkarar-Gradwohl et al., 2014), and French (Pahlavan, Mouchiroud, Zenasni, & Panksepp, 2008) samples. Their congruent findings showed positive correlations between PLAYFULNESS and Extraversion, CARING and Agreeableness, SEEKING and Openness to Experience, as well as negative correlations between FEAR, ANGER, and SADNESS and Emotional Stability.

Despite the growing literature about the ANPS, further studies are needed to explore its psychometric properties in more detail and to determine its appropriate use in both research and clinical practice. Its psychometric properties have been studied in various languages and samples: United States English (Davis & Panksepp, 2011), French (Pingault, Pouga, et al., 2012), Spanish (Abella, Panksepp, Manga, Bárcena, & Iglesias, 2011), Italian (Pascazio et al., 2015), and Norwegian (Geir, Selsbakk, Theresa, & Sigmund, 2014). These studies identified several strengths but also noted psychometric properties that could be improved (Pingault, Falissard, Côté, & Berthoz, 2012). Moreover, its length (14 items per subscale, for a total of 84 items) raises questions about its practicality in surveys or longitudinal studies in which numerous questionnaires are administered. A short version of the French ANPS (ANPS-S) has therefore been developed (Pingault, Falissard, et al., 2012), composed of 36 items from the original items (6 for each scale). The validation of the ANPS-S in both French (N=830) and Canadian French (N=431) samples showed improved psychometric properties. This short version is different from that of Barrett and coll. (Barrett, Robins, & Janata, 2013), which included several new items not in the long version.

Although more than 10 papers have been published on the psychometric properties of the ANPS (short or long versions), no study has yet investigated the measurement invariance of this instrument. Measurement invariance assesses whether scales measure the same construct regardless of the group or the occasion of measurement (the latter is known as longitudinal invariance). Unless a scale is known to be invariant, we cannot determine if the observed score difference between two groups or two waves of measurement is due to a real difference or to changes in the structure of the construct across groups or times of assessment (Brown, 2006). For example, for a statistically significant difference in the mean score to a questionnaire between men and women to be trusted to reveal sex differences, men and women must have a similar understanding of the items evaluating the latent trait. In addition, because these are supposed to measure temperamental or personality characteristics (conceptualized as stable over time), longitudinal invariance is required to evaluate long-term stability. Therefore, measurement invariance is essential to appropriately assess between-group differences or temporal changes in a construct.

This study sought for the first time to assess (i) the longitudinal measurement invariance and the long-term stability of the ANPS, and (ii) the sex measurement invariance of the ANPS in a large sample of Canadian families who were followed longitudinally.

Methods

Sample

The study sample comprises participants in the EMIGARDE cohort (Côté et al., 2013), a longitudinal study of child development conducted in Montreal (Quebec, Canada) from 2003 to 2011 with 4 collection waves (2004-2005-2006-2010). The initial sample was composed of 499 families assessed by several measures concerning both the children and their parents. Parents completed the ANPS long version at the third (2006, hereafter T1) and fourth

(2010, hereafter T2) data collection waves for personality assessments. Specifically, a subgroup of 520 subjects completed it at T1, and 569 at T2. After we excluded questionnaires for which more than 10% of ANPS items were missing (N=11 at T1 and N=1 at T2), the final sample included 509 subjects (222 men and 287 women) at T1 and 568 (249 men and 319 women) at T2, with data at both time points for 422 subjects (177 men and 245 women). The mean age of the participants at T1 was 36.5±5.8 years; on average, the men were 3 years older than the women (38.4±6.3 versus 35.2±5.0). Most participants had intermediate to high levels of education: 56.4% had a university degree, 24.7% had graduated from high school, 8.6% had some college education, 7.9% some high school, and only 2.4% had no secondary education.

Measure

We used the French adaptation of the Affective Neuroscience Personality Scales, ANPS version 2.4 (Pahlavan, Mouchiroud, Zenasni, & Panksepp, 2008) – hereafter referred to as the ANPS long version, ANPS-L. Besides the six emotional subscales, the original ANPS included a SPIRITUALITY subscale, which was not based on neuro-ethological models and which we chose not to include in our survey. Each ANPS-L scale comprised 14 items. Items were answered on a 4-point scale ranging from “I totally disagree” to “I totally agree”. As described above, the ANPS-Short (ANPS-S) version includes a selection of 36 items from the original items (6 for each scale), and the ANPS-S subscale scores used in these analyses were computed from the participants’ responses to the ANPS-L. The internal consistency of each ANPS dimension was assessed by an ordinal version of Cronbach’s alpha, which takes into account the ordinal nature of the items; it is calculated with the polychoric correlation matrix instead of the usual Pearson correlation matrix (Gadermann, Guhn, & Zumbo, 2012; Zumbo, Gadermann, & Zeisser, 2007). Values lower than .70 were considered unsatisfactory, between

.70 and .79 fair, between .80 and .89 good, and ≥.90 excellent (Cicchetti, 1994).

Assessment of the measurement invariance

Both longitudinal invariance and sex invariance were tested with Confirmatory Factor Analysis (CFA) models and a weighted least squares means- and variance-adjusted estimator (WLSMV) with Theta parameterization to take into account the ordinal nature of ANPS items. Longitudinal invariance was assessed with single-group CFA where the latent factors as well as the residuals for each item were allowed to correlate between T1 and T2. Sex invariance was studied at both time points with multiple-group CFA (MGCFA) that compared the factor structure across sex (Brown, 2006; Gregorich, 2006; Kline, 2010; Millsap, 2011).

The sequence of models for testing measurement invariance varies widely between studies (Schmitt & Kuljanin, 2008; Vandenberg & Lance, 2000). Of the 13 models proposed by Marsh (Marsh, Morin, Parker, & Kaur, 2014), we consecutively tested five levels of invariance, corresponding to five nested models with increasing constraints. For sex invariance, the same model (**Figure 1**) was hypothesized in both groups. In the model list below, names in square brackets correspond to common alternative terminology for these models; Greek letters refer to parameters in **Figure 1**; see **Table S1** for the details of the model parameterization:

1. Configural invariance (unconstrained factor loadings [λ], same subset of items associated with the same construct);
2. Metric invariance [weak factorial] (equal factor loadings [λ] across times for longitudinal invariance or groups for sex invariance);
3. Scalar invariance [strong factorial] (equal factor loadings [λ] and item thresholds [τ]);
4. Residual invariance [strict factorial] (equal factor loadings [λ], item thresholds [τ], and item residual variances [ε]);

5. Structural invariance [complete factorial] (equal factor loadings [λ], item thresholds [τ], item residual variance [ε], factor variance-covariances [ϕ], and factor means [ξ]).

We followed the same sequence for longitudinal invariance, hypothesizing the same model (**Figure 1**) for both waves, with the constraints set consecutively across waves.

Configural invariance was evaluated with three model-fit indices: the Chi-square test (highly affected by sample size), the Comparative Fit Index (CFI, acceptable fit if >0.95, poor fit if <0.90, otherwise marginal) and the Root Means Square Error Approximation (RMSEA, acceptable fit if <0.06) (Hu & Bentler, 1999). Then, if the difference in the fit indices (ΔCFI and ΔRMSEA) between a model and the (preceding) less constrained model was equal or less than –0.01 for ΔCFI and equal or less than 0.015 for ΔRMSEA, we considered that the level of measurement invariance was achieved (Chen, 2007; Cheung & Rensvold, 2002; Marsh, Nagengast, & Morin, 2013). Although these criteria are those used most commonly in the measurement invariance literature, Meade et al. (2008) have proposed more stringent criteria (i.e., cutoff of ΔCFI>0.002 to define violation of invariance). As they noted (Meade et al. 2008), however, researchers must exercise their judgment in these situations: there is a difference between detectable non-invariance (relevant from a methodological perspective) and practically significant non-invariance (relevant from an empirical perspective; Nye & Drasgow, 2011). In particular, the ΔCFI cutoff of 0.002 may be useful for the first aim, but less useful for the second. We therefore chose the cutoff of ΔCFI -0.01 and ΔRMSEA 0.015 in our study. The nested Chi-square test between two models (robust chi-square-based likelihood ratio adjusted for means and variance, DIFFTEST in Mplus, Muthén & Muthén, 1998-2010) was not used because of its recognized sensitivity to sample size, whereas ΔCFI is independent of both the model's sample size and its overall CFI (Cheung & Rensvold, 2002).

Partial invariance

When we found that the model's goodness of fit worsened substantially (i.e. ΔCFI>-.01), we identified the non-invariant item(s) by reviewing the modification indices and then removed the corresponding equality constraint between the two groups (or waves) (i.e., the parameter was freely estimated in each group or at each time). If the differences between the CFIs and RMSEAs in the resulting and the less constrained models exceeded the accepted cutoffs, partial invariance was achieved, and the parameter remained unconstrained in the subsequent models of the measurement invariance assessment process.

Long-term stability

We assessed the stability of the measure over time with Intraclass Correlation Coefficients (ICC; consistency version, corresponding to a one-way random effects ANOVA model, or ICC [1,1] in Shrout & Fleiss, 1979). As recommended by Cicchetti (1994), we classified ICC values as follows: ICC>.75 excellent, from .60 to .74 good, .40 to .59 fair, and .40 poor. We used the bootstrap procedure to calculate their 95% confidence intervals.

Software

R version 3.0 (R Core Team, 2013) was used for data management, descriptive analyses, Cronbach's alphas, and ICC analyses, and Mplus version 7 (Muthén & Muthén, 1998-2010) for CFA.

Results

Descriptive statistics

Data were missing for a few items at both T1 and T2 (**Table S2**). **Table 1** (T1) and **Table 2** (T2) report the scores of the six dimensions of the ANPS-L and the ANPS-S. The internal consistency of the long version was fair to excellent for all scales at both time points (Cronbach's alpha range: .76-.90). For the short version, it was fair to good for 4 scales (range .75-.84) and slightly lower (.67-.69) for the other two (CARING and PLAYFULNESS) at both time points. This difference between the long and short versions was expected because the number of items influences Cronbach's alpha. The implementation of a recent adaptation of Cronbach's alpha to ordinal items (Gadermann, Guhn, & Zumbo, 2012; Zumbo, Gadermann, & Zeisser, 2007) yielded substantially improved estimates of internal reliability compared with previous estimates (Pingault, Falissard, Côté, & Berthoz, 2012).

Longitudinal invariance of the ANPS-L and ANPS-S

The results for the analysis of longitudinal invariance are reported in **Table 3**. For both the long and short versions of the ANPS, the fit of the configural model was acceptable according to the RMSEA (<0.06) but poor according to CFI (<0.90). The differences in the CFI and RMSEA were below the accepted cutoffs for both versions at each step of the measurement invariance assessment process; full longitudinal invariance was thus demonstrated for the ANPS-L and the ANPS-S.

Long-term stability of the ANPS-L and ANPS-S

The stability of the scores at T1 and T2 were assessed with the ICCs, reported in **Table 4**. The ICCs of both the long and short versions of the ANPS were similar (overlapping 95% CIs) for all dimensions, as were those for men and women.

For the long version, the ICCs for SEEKING and SADNESS both had ICC values classified as good (i.e., between .60 and .74), and the values for the other four were excellent (>.75). The ICCs values for the short version nearly all fell in the good range, except that for SADNESS, which was fair ICC (.40< ICC<.60; Cicchetti, 1994).

Measurement invariance across sex for the ANPS-L and ANPS-S

Table 5 summarizes the goodness-of-fit indices for measurement invariance across sex for the ANPS-L. The configural model showed a good fit according to the RMSEA (.034; 90%CI .031–.036) although the CFI was below the most commonly accepted threshold (CFI=.812).

When we applied the different levels of constraint, the CFI did not worsen substantially when we assessed metric and scalar invariance. When residual invariance was assessed, however, the decreased in CFI of .010 indicated a lack of invariance. Partial residual invariance was achieved, however, when we allowed the residual of the item Anger 6 ("When I am frustrated, I rarely become angry") to be freely estimated in one group. In the following step, we could not establish the partial complete invariance (i.e., that means and variance-covariance matrices were equal across groups). Modification indices suggested that model fit would have been improved by freeing the means of the following factors: CARING, FEAR, ANGER, and SADNESS.

Results were similar for the ANPS-S (**Table 5**). Acceptable fit indices (CFI=.919 and RMSEA=.040 [90%CI .035–.044]) were found, hence we showed configural invariance. The model then showed metric invariance (ΔCFI=.000, ΔRMSEA=-.001), but failed to show full scalar invariance according to CFI (ΔCFI=-.010). We could, however, obtain partial invariance by releasing only one threshold from the equality constraint (the second threshold of the item PLAYFULNESS 11 "I like all kinds of games including those with physical contact"). Residual invariance was shown by the acceptable decrease of CFI and RMSEA when we constrained item residual variances to equality. Finally, consistently with findings for the long version, we failed to establish complete invariance. Modification indices suggested that the equality constraint should be released for the means of the same factors as for the long version.

After showing scalar measurement invariance (except from one threshold), we compared statistically the means of the scores for men and women. Significant sex differences were found for 5 of the 6 dimensions (**Table 1** and **Table 2**). At T1, women reported higher CARING, FEAR, ANGER, and SADNESS scores, and lower PLAYFULNESS scores. At T2, the pattern remained almost the same, except that sex differences for ANGER were not statistically significant (ANPS-L).

To summarize, the long version of the ANPS showed full measurement invariance across sex at the scalar level, and partial measurement invariance (residual variance was non-invariant for one item) at the residual level. The short versions of the ANPS showed full metric invariance across sex, and partial scalar and residual invariance (one threshold was non-invariant). Neither the long nor short version showed complete invariance.

Discussion

The aims of this psychometric study were to investigate (i) the longitudinal measurement invariance and long-term stability, and (ii) the sex measurement invariance of the ANPS (both the long and short French versions) and sex differences.

Measurement invariance is a prerequisite for meaningful comparisons across groups or time points, and lack of invariance can lead to misleading interpretations of change scores and group differences. Comparisons of group means are based on the assumption of measurement invariance, but this is rarely tested empirically.

Longitudinal properties of the ANPS

In this study we found that both the long and short versions of the ANPS had full longitudinal invariance.

Longitudinal invariance was ascertained at the level of both the measurement model (i.e., the same subset of items associated with the same construct, their item loadings, item thresholds, and residuals did not vary significantly over time) and the instrument structure (i.e. means of the factors, variance and covariance of the latent factors). The first can be sufficient to establish comparisons of mean scores over time. Some authors (Marsh et al., 2013) have also suggested that in cases of multifactorial constructs with meaningful associations between latent factors (e.g., for establishing personality profiles), changes in the relations between latent factors over time might be cause for concern. Our findings thus strongly support the conclusion that the ANPS measures a personality trait (i.e., is stable over time).

Furthermore, we also showed that the ANPS has good long-term stability: all dimensions of the long version have good to excellent ICCs (varying from .67 to .78), and all but one dimension (SADNESS) of the short version had good ICCs (varying from to .59 to .74).

These results imply that the construct measured by the ANPS is stable and reliable over time. Since this is the first study to demonstrate this measure's stability across time, our findings, which indicate that the ANPS measures emotional-based personality traits and not emotional states, need to be replicated.

Across-sex properties of the ANPS

We showed full scalar sex invariance for the ANPS-L and partial scalar sex invariance for the ANPS-S. Partial scalar invariance was obtained by releasing only one threshold from the constraint for equality across sexes. Although there is no agreement about an acceptable level of partial invariance, we think that one threshold of 108 can be considered a negligible deviation from full invariance. These findings suggest that the observed sex score differences are representative of differences on the latent factors of the ANPS (for both the long and short versions). Therefore, they can be reliably interpreted as actual differences in the latent constructs representing these dimensions (Cheung & Rensvold, 2002).

Partial residual invariance was also obtained for both ANPS-L and ANPS-S, as non-invariant residual variance can be considered negligible if it concerns only one of 36 items in the ANPS-S or one of 84 in the ANPS-L. Residual invariance indicates that "for both groups, items have the same quality as measures of the underlying construct" (Cheung & Rensvold, 2002, p 236). Although achieving measurement invariance at this level shows that the items have equivalent properties across sex, residual invariance is not mandatory for between-group comparisons.

Finally, significant mean differences were found for 5 dimensions of the ANPS-L (FEAR, ANGER, SADNESS, CARING, and PLAYFULNESS) and in 4 dimensions of the ANPS-S (FEAR, ANGER, SADNESS, and CARING). Consistently, these dimensions were those that needed to be released from the mean equality constraint in our complete measurement invariance models to achieve invariance.

As expected, the mean differences observed in this study are similar to those reported in other studies of the ANPS (Davis, Panksepp, & Normansell, 2003; Geir, Selsbakk, Theresa, & Sigmund, 2014; Pahlavan, Mouchiroud, Zenasni, & Panksepp, 2008; Pingault, Pouga, Grèzes, & Berthoz, 2012). These differences are also consistent with other studies showing a greater propensity for nurturing (Davis, Panksepp, & Normansell, 2003; Derntl et al., 2010) and a higher prevalence of depressive and anxious feelings (McLean & Anderson, 2009; Parker & Brotchie, 2010) among women.

Comparisons between the long and short versions of the ANPS

As expected, in both longitudinal and across-sex invariance models, model fit was significantly better in the short than in the version of the ANPS. The two versions showed similar ICC values and thus similar long-term stability, with overlapping 95% CIs for each scale. Thus these findings offer further validation of the good psychometric properties of the ANPS short version.

Strengths and limitations

The size of our sample was adequate for our research questions and is representative of the population from which it was selected. However, some limitations should be considered.

The first concerns the generalizability of our findings. Our sample is mostly composed of educated parents of young children from Montreal (Canada). Further studies should examine if these results remain the same in other populations that differ in age, culture, level of education, or socioeconomic status. In particular, it might be interesting to investigate the cultural invariance of the ANPS. Studies using the FFM have showed that personality traits vary across culture (Costa, Terracciano, & McCrae, 2001), hence addressing this issue with the ANPS – which are not based on the same lexical approach – may prove interesting.

Second, only the ANPS-L was administered to our sample, and the ANPS-S was derived from the items of the ANPS-L. In questionnaire surveys, respondents tend to give faster and more uniform answers in the last part of the questionnaire (Galesic & Bosnjak, 2009). Our results might therefore have been different had the ANPS-S been administered directly.

Furthermore, our participants belong to the general population, and it would be interesting to investigate the psychometric properties of the ANPS within clinical groups. Some studies have used this instrument in clinical populations: the first were conducted by Savitz and colleagues among South African patients diagnosed with affective disorders (Savitz, Van der Merwe, & Ramesar, 2008a; J. Savitz, Van Der Merwe, & Ramesar, 2008b); another by Geir and colleagues among Norwegian patients diagnosed with personality disorders (Geir et al., 2014), and still another by Carré and colleagues among adults with an Autism spectrum condition (Carré et al., 2015).

A final methodological remark concerns the less than optimal fit of some of our configural models, according to the fit indices we report here. This may create concerns for the global adequacy of the ANPS. However, three points should be considered. First, although model fit was sometimes not adequate according to the CFI, all our configural models showed good fit according to the RMSEA. Second, it is well known in the literature that personality measures (such as NEO-Personality Inventory and Big Five Inventory) suffer from low fit indices (in particular, CFI) and often fail to demonstrate adequate model fit in confirmatory factor analysis studies (Booth & Hughes, 2014). This issue is due mainly to the presence of cross-loadings, which are not allowed in CFA. Some authors (Marsh, Morin, Parker, & Kaur, 2014) have thus proposed the use of Exploratory Structural Equation Models (ESEM) to evaluate the fit of personality instruments. ESEM enables all items to load on each factor (arguing that zero cross-loadings is an excessively restrictive hypothesis), and the only a-priori assumption is the number of factors. As a consequence of these different specifications, ESEM yield better fit indices. However, this vision is not unanimously shared (Booth & Hughes, 2014), mainly because ESEM is an exploratory tool and modeling all possible cross-loadings contradicts the principle of parsimony. We agree with these arguments and thus chose a CFA framework for this study, even though it came at the price of lower CFI values. Third, in this study we were interested in evaluating measurement invariance. According to Marsh et al., the cutoff values for goodness-of-fit indices represent only rough guidelines, and "it is typically more useful to compare the relative fit of different models in a nested or partially nested taxonomy of models designed a priori to evaluate particular aspects of interest than to compare the relative fit of single models" (Marsh et al., 2013, p. 1220). Finally, the structural properties of the ANPS have been studied and discussed in previous papers (Barrett et al., 2013; Pingault, Falissard, et al., 2012; Pingault, Pouga, et al., 2012).

Despite these limitations, this is the first study demonstrating longitudinal and sex invariance as well as long-term stability for the ANPS and presenting Cronbach alphas that take the ordinal nature of the items into account. These results thus add to the extant literature

References

Abella, V., Panksepp, J., Manga, D., Bárcena, C., & Iglesias, J. A. (2011). Spanish validation of the Affective Neuroscience Personality Scales. *The Spanish Journal of Psychology*, *14*(2), 926–935.

Barrett, F. S., Robins, R. W., & Janata, P. (2013). A brief form of the Affective Neuroscience Personality Scales. *Psychological Assessment*, *25*(3), 826–843. http://doi.org/10.1037/a0032576

Berthoz, S., Orvoën, H., & Grezes, J. (2010). High temperamental fearfulness and reduced right amygdala volume: An emotional endophenotype? Presented at the Sixth Annual Meeting of the European Psychiatric Association, Paris, France.

Booth, T., & Hughes, D. J. (2014). Exploratory structural equation modeling of personality data. *Assessment*, *21*(3), 260–271. http://doi.org/10.1177/1073191114528029

Brown, T. (2006). *Confirmatory Factor Analysis for Applied Research, First Edition* (1 edition). New York: The Guilford Press.

Carré, A., Chevallier, C., Robel, L., Barry, C., Maria, A.-S., Pouga, L., … Berthoz, S. (2015). Tracking Social Motivation Systems Deficits: The Affective Neuroscience View of Autism. *Journal of Autism and Developmental Disorders*. http://doi.org/10.1007/s10803-015-2498-2

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. http://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. http://doi.org/10.1207/S15328007SEM0902_5

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284−290. http://doi.org/10.1037/1040-3590.6.4.284

Cloninger, C. R. (1987). Neurogenetic adaptive mechanisms in alcoholism. *Science*, *236*(4800), 410−416. http://doi.org/10.1126/science.2882604

Cloninger C, Svrakic DM, & Przybeck TR. (1993). A psychobiological model of temperament and character. *Archives of General Psychiatry*, *50*(12), 975−990. http://doi.org/10.1001/archpsyc.1993.01820240059008

Costa, P. J., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, *81*(2), 322−331. http://doi.org/10.1037/0022-3514.81.2.322

Côté, S. M., Mongeau, C., Japel, C., Xu, Q., Séguin, J. R., & Tremblay, R. E. (2013). Child care quality and cognitive development: trajectories leading to better preacademic skills. *Child Development*, *84*(2), 752−766. http://doi.org/10.1111/cdev.12007

Davis, K. L., & Panksepp, J. (2011). The brain’s emotional foundations of human personality and the Affective Neuroscience Personality Scales. *Neuroscience and Biobehavioral Reviews*, *35*(9), 1946−1958. http://doi.org/10.1016/j.neubiorev.2011.04.004

Davis, K. L., Panksepp, J., & Normansell, L. (2003). The Affective Neuroscience Personality Scales: Normative Data and Implications. *Neuro-Psychoanalysis*, *5*(1), 57−69. http://doi.org/10.1080/15294145.2003.10773410

Derntl, B., Finkelmeyer, A., Eickhoff, S., Kellermann, T., Falkenberg, D. I., Schneider, F., & Habel, U. (2010). Multidimensional assessment of empathic abilities: neural correlates and gender differences. *Psychoneuroendocrinology*, *35*(1), 67−82. http://doi.org/10.1016/j.psyneuen.2009.10.006

Ersche, K. D., Turton, A. J., Pradhan, S., Bullmore, E. T., & Robbins, T. W. (2010). Drug Addiction Endophenotypes: Impulsive Versus Sensation-Seeking Personality Traits. *Biological Psychiatry*, *68*(8), 770−773. http://doi.org/10.1016/j.biopsych.2010.06.015

Farinelli, M., Panksepp, J., Gestieri, L., Leo, M. R., Agati, R., Maffei, M., … Northoff, G. (2013). SEEKING and depression in stroke patients: an exploratory study. *Journal of Clinical and Experimental Neuropsychology*, *35*(4), 348−358. http://doi.org/10.1080/13803395.2013.776009

Farinelli, M., Panksepp, J., Gestieri, L., Maffei, M., Agati, R., Cevolani, D., … Northoff, G. (2015). Do brain lesions in stroke affect basic emotions and attachment? *Journal of Clinical and Experimental Neuropsychology*, *37*(6), 595−613. http://doi.org/10.1080/13803395.2014.991279

Felten, A., Montag, C., Markett, S., Walter, N. T., & Reuter, M. (2011). Genetically determined dopamine availability predicts disposition for depression. *Brain and Behavior*, *1*(2), 109−118. http://doi.org/10.1002/brb3.20

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating Ordinal Reliability for Likert-Type and Ordinal Item Response Data: A Conceptual, Empirical, and Practical Guide. *Practical Assessment, Research & Evaluation*, *17*(3).

Galesic, M., & Bosnjak, M. (2009). Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly*, *73*(2), 349−360. http://doi.org/10.1093/poq/nfp031

Geir, P., Selsbakk, J. M., Theresa, W., & Sigmund, K. (2014). Testing Different Versions of the Affective Neuroscience Personality Scales in a Clinical Sample. *PLoS ONE*, *9*(10). http://doi.org/10.1371/journal.pone.0109394

Gottesman, I. I., & Gould, T. D. (2003). The endophenotype concept in psychiatry: etymology and strategic intentions. *The American Journal of Psychiatry*, *160*(4), 636−645. http://doi.org/10.1176/appi.ajp.160.4.636

Gray, J. A. (1987). *The Psychology of Fear and Stress*. CUP Archive.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, *44*(11 Suppl 3), S78−94. http://doi.org/10.1097/01.mlr.0000245454.12228.8f

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

John, O. P., & Srivastava, S. (2001). The big five trait taxonomy: history, measurement, and theoretical perspectives. In *Handbook of personality. Theory and research* (Pervin, LA & John, OP). New York: Guilford Press.

Kelley, W. M., Wagner, D. D., & Heatherton, T. F. (2015). In Search of a Human Self-Regulation System. *Annual Review of Neuroscience*, *38*, 389−411. http://doi.org/10.1146/annurev-neuro-071013-014243

Kendler, K. S., Zachar, P., & Craver, C. (2011). What kinds of things are psychiatric disorders? *Psychological Medicine*, *41*(6), 1143−1150. http://doi.org/10.1017/S0033291710001844

Kline, R. (2010). *Principles and Practice of Structural Equation Modeling, Third Edition* (3rd edition). New York: The Guilford Press.

Kret, M. E., & Ploeger, A. (2015). Emotion processing deficits: a liability spectrum providing insight into comorbidity of mental disorders. *Neuroscience and Biobehavioral Reviews*, *52*, 153−171. http://doi.org/10.1016/j.neubiorev.2015.02.011

Krueger, R. F., & Eaton, N. R. (2015). Transdiagnostic factors of mental disorders. *World Psychiatry*, *14*(1), 27−29. http://doi.org/10.1002/wps.20175

Krueger, R. F., & Markon, K. E. (2014). The role of the DSM-5 personality trait model in moving toward a quantitative and empirically based approach to classifying personality and psychopathology. *Annual Review of Clinical Psychology*, *10*, 477−501. http://doi.org/10.1146/annurev-clinpsy-032813-153732

Maj, M. (2014). Keeping an open attitude towards the RDoC project. *World Psychiatry*, *13*(1), 1−3. http://doi.org/10.1002/wps.20111

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014a). Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*, 85−110. http://doi.org/10.1146/annurev-clinpsy-032813-153700

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014b). Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*, 85−110. http://doi.org/10.1146/annurev-clinpsy-032813-153700

Marsh, H. W., Nagengast, B., & Morin, A. J. S. (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental Psychology*, *49*(6), 1194−1218. http://doi.org/10.1037/a0026913

McGorry P, & Nelson B. (2016). WHy we need a transdiagnostic staging approach to emerging psychopathology, early diagnosis, and treatment. *JAMA Psychiatry*, 1−2. http://doi.org/10.1001/jamapsychiatry.2015.2868

McGorry, P., & van Os, J. (2013). Redeeming diagnosis in psychiatry: timing versus specificity. *Lancet (London, England)*, *381*(9863), 343−345. http://doi.org/10.1016/S0140-6736(12)61268-9

McLean, C. P., & Anderson, E. R. (2009). Brave men and timid women? A review of the gender differences in fear and anxiety. *Clinical Psychology Review*, *29*(6), 496−505. http://doi.org/10.1016/j.cpr.2009.05.003

Millsap, R. E. (2011). *Statistical Approaches to Measurement Invariance* (1 edition). New York: Routledge.

Montag, C., & Reuter, M. (2014). Disentangling the molecular genetic basis of personality: From monoamines to neuropeptides. *Neuroscience & Biobehavioral Reviews*, *43*, 228−239. http://doi.org/10.1016/j.neubiorev.2014.04.006

Montag, C., Reuter, M., Jurkiewicz, M., Markett, S., & Panksepp, J. (2013). Imaging the structure of the human anxious brain: a review of findings from neuroscientific personality psychology. *Reviews in the Neurosciences*, *24*(2), 167−190. http://doi.org/10.1515/revneuro-2012-0085

Muthén, L., & Muthén, B. (1998-2010). *Mplus User's Guide. Sixth Edition.* Los Angeles, CA: Muthén & Muthén.

Özkarar-Gradwohl, F. G., Panksepp, J., İçöz, F. J., Çetinkaya, H., Köksal, F., Davis, K. L., & Scherler, N. (2014). The influence of culture on basic affective systems: the comparison of Turkish and American norms on the affective neuroscience personality scales. *Culture and Brain*, *2*(2), 173−192. http://doi.org/10.1007/s40167-014-0021-9

Pahlavan, F., Mouchiroud, C., Zenasni, F., & Panksepp, J. (2008). Validation de l'adaptation française de l'échelle neuro-affective de personnalité. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, *58*(3), 155−163. http://doi.org/10.1016/j.erap.2007.08.004

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford: Oxford University Press.

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, *14*(1), 30−80. http://doi.org/10.1016/j.concog.2004.10.004

Panksepp, J. (2006). Emotional endophenotypes in evolutionary psychiatry. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *30*(5), 774−784. http://doi.org/10.1016/j.pnpbp.2006.01.004

Panksepp, J., & Biven, L. (2012). *The Archaeology of Mind: Neuroevolutionary Origins of Human Emotions* (1 edition). New York: W. W. Norton & Company.

Parker, G., & Brotchie, H. (2010). Gender differences in depression. *International Review of Psychiatry (Abingdon, England)*, *22*(5), 429−436. http://doi.org/10.3109/09540261.2010.492391

Pascazio, L., Bembich, S., Nardone, I. B., Vecchiet, C., Guarino, G., & Clarici, A. (2015). Validation of the italian translation of the affective neuroscience personality scales,. *Psychological Reports*, *116*(1), 97−115. http://doi.org/10.2466/08.09.PR0.116k13w4

Pingault, J.-B., Falissard, B., Côté, S., & Berthoz, S. (2012). A New Approach of Personality and Psychiatric Disorders: A Short Version of the Affective Neuroscience Personality Scales. *PLoS ONE*, *7*(7), e41489. http://doi.org/10.1371/journal.pone.0041489

Pingault, J.-B., Pouga, L., Grèzes, J., & Berthoz, S. (2012). Determination of emotional endophenotypes: A validation of the Affective Neuroscience Personality Scales and further perspectives. *Psychological Assessment*, *24*(2), 375−385. http://doi.org/10.1037/a0025692

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

Reuter, M., Weber, B., Fiebach, C. J., Elger, C., & Montag, C. (2009). The biological basis of anger: Associations with the gene coding for DARPP-32 (PPP1R1B) and with amygdala volume. *Behavioural Brain Research*, *202*(2), 179−183. http://doi.org/10.1016/j.bbr.2009.03.032

Saucier, G. (2009). Recurrent Personality Dimensions in Inclusive Lexical Studies: Indications for a Big Six Structure. *Journal of Personality*, *77*(5), 1577−1614. http://doi.org/10.1111/j.1467-6494.2009.00593.x

Savitz, J., Van Der Merwe, L., & Ramesar, R. (2008). Hypomanic, cyclothymic and hostile personality traits in bipolar spectrum illness: a family-based study. *Journal of Psychiatric Research*, *42*(11), 920−929. http://doi.org/10.1016/j.jpsychires.2007.10.011

Savitz, J., Van Der Merwe, L., & Ramesar, R. (2008). Personality endophenotypes for bipolar affective disorder: a family-based genetic association analysis. *Genes, Brain and Behavior*, *7*(8), 869−876. http://doi.org/10.1111/j.1601-183X.2008.00426.x

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210−222. http://doi.org/10.1016/j.hrmr.2008.03.003

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420−428.

Siever, L. J., & Weinstein, L. N. (2009). The neurobiology of personality disorders: implications for psychoanalysis. *Journal of the American Psychoanalytic Association*, *57*(2), 361−398. http://doi.org/10.1177/0003065109333502

Stein, M. B., & Stein, D. J. (2008). Social anxiety disorder. *Lancet (London, England)*, *371*(9618), 1115−1125. http://doi.org/10.1016/S0140-6736(08)60488-2

Toronchuk, J. A., & Ellis, G. F. R. (2013). Affective Neuronal Selection: The Nature of the Primordial Emotion Systems. *Frontiers in Psychology*, *3*. http://doi.org/10.3389/fpsyg.2012.00589

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

Zumbo, B., Gadermann, A., & Zeisser, C. (2007). Ordinal Versions of Coefficients Alpha and Theta for Likert Rating Scales. *Journal of Modern Applied Statistical Methods*, *6*(1).

Figure 1. Hypothesized model tested in the Confirmatory Factor Analysis

The figure represents the hypothesized CFA model of the ANPS-S (the same holds true for the ANPS-L). Ellipses represent unobserved latent factors, rectangles observed variables, single-headed arrows the impact of one variable on another, and double-headed arrows correlations between pairs of variables. The configural invariance model tested whether the fit of the hypothesized model is acceptable in both groups without parameter constraints. Testing metric invariance allowed us to evaluate the model fit when the magnitude of the loadings (λ_i) was fixed equal across sex. Scalar invariance was tested by adding the additional constraints of item thresholds equality across sex (eg, each of the 3 thresholds of item 1 "$\tau_{\text{Item 1 SEEK}}$" in men equal to "$\tau_{\text{Item 1 SEEK}}$" in women). When residual invariance was tested, residual variances (ε_i) were forced to be equal in both groups, to determine whether the unexplained part of the model (i.e., the error terms in the regression equations) was the same in both groups. Finally, when complete invariance was tested, the factor variances (ϕ_i), covariances ($\phi_{i,j}$), and means (ξ_i) were constrained to be equal across groups. The hypothesized model was the same for the longitudinal measurement invariance, except that (i) residual correlations between the same items at T1 and T2 were set, and (ii) a single-group CFA was used rather than a multiple-groups CFA.

Table 1. Descriptive statistics of the six dimensions of the ANPS-L and ANPS-S at Time 1

	Cronbach alpha	*Total sample*			*Men*			*Women*			*Sex diff (g)*
		Mean (SD)	*Median (IQR)*	*Min-Max*	*Mean (SD)*	*Median (IQR)*	*Min-Max*	*Mean (SD)*	*Median (IQR)*	*Min-Max*	
Long version											
SEEKING	.77	27.9 (5.07)	28 (25-31)	(6-42)	27.83 (5.19)	27.5 (25-31)	(8-42)	27.96 (4.98)	28 (25-31)	(6-39)	.03
CARING	.79	26.93 (5.68)	27 (23-31)	(10-40)	25.06 (5.65)	25 (22-29)	(10-39)	28.36 (5.28)	28 (25-32)	(16-40)	.63***
PLAYFULNESS	.82	27.38 (5.77)	28 (24-32)	(9-41)	28.02 (5.88)	28.5 (24-32)	(9-41)	26.88 (5.64)	27 (23-31)	(10-41)	-.20*
FEAR	.89	19.25 (7.08)	19 (15-23)	(2-40)	17.26 (6.62)	17 (13-21)	(4-39)	20.78 (7.06)	21 (16-26)	(2-40)	.51***
ANGER	.84	15.94 (6.08)	16 (12-20)	(1-33)	15.21 (6.1)	15 (11-19)	(1-33)	16.49 (6.02)	16 (12-20)	(3-33)	.21*
SADNESS	.80	18.78 (5.75)	18.5 (15-22)	(3-38)	16.82 (5.4)	16 (14-20)	(3-32)	20.3 (5.57)	20 (16.75-24)	(8-38)	.67***
Short version											
SEEKING	.75	12.77 (2.95)	13 (11-15)	(1-18)	12.85 (3.03)	13 (11-15)	(4-18)	12.7 (2.89)	13 (11-15)	(1-18)	.05
CARING	.68	12.24 (2.68)	12 (11-14)	(4-18)	11.61 (2.65)	12 (10-14)	(4-17)	12.73 (2.6)	13 (11-15)	(5-18)	.54***
PLAYFULNESS	.67	12.29 (2.78)	12 (10-14)	(5-18)	12.46 (2.78)	13 (11-14)	(3-18)	12.23 (2.78)	12 (10-14)	(5-17)	-.08
FEAR	.82	7.57 (3.39)	7 (6-10)	(0-18)	6.4 (3.1)	7 (4-8)	(0-17)	8.47 (3.34)	8 (6-11)	(0-18)	.64***
ANGER	.81	7.23 (3.35)	7 (5-9)	(0-17)	6.85 (3.4)	7 (4-9)	(0-17)	7.51 (3.28)	7 (5-9.25)	(0-17)	.20*
SADNESS	.77	6.07 (3.11)	6 (4-8)	(0-16)	5.26 (2.95)	5 (3-7)	(0-16)	6.69 (3.09)	6 (5-8)	(0-16)	.47***

The table presents the descriptive statistics – mean and standard deviation (SD), median and interquartile range (IQR), minimum and maximum – for the total sample, by sex, and, by ANPS version. The first column reports Cronbach's alpha (version for ordinal items). The last column indicates Hedges' g (effect size) for the differences between men and women. For each dimension, the range of possible scores is 0-42 for the ANPS-L and 0-18 for the ANPS-S. P-values refer to the t-test: *=p<.050; **=p<.010; *=p<.001

Table 2. Descriptive statistics of the six dimensions of the ANPS-L and ANPS-S at Time 2

	Cronbach alpha	*Total sample*			*Men*			*Women*			*Sex diff (g)*
		Mean (SD)	*Median (IQR)*	*Min-Max*	*Mean (SD)*	*Median (IQR)*	*Min-Max*	*Mean (SD)*	*Median (IQR)*	*Min-Max*	
Long version											
SEEKING	.77	27.82 (4.78)	28 (25-31)	(9-41)	27.6 (4.79)	28 (25-31)	(11-41)	28 (4.77)	28 (25-31)	(9-40)	.08
CARING	.79	27 (5.61)	27 (24-31)	(8-42)	24.84 (5.46)	25 (21-28)	(8-40)	28.75 (5.11)	29 (25-32)	(14-42)	.74***
PLAYFULNESS	.82	26.82 (5.52)	27 (23-30)	(9-42)	27.46 (5.56)	27 (24-31.75)	(14-42)	26.31 (5.45)	26 (23-30)	(9-41)	-.22*
FEAR	.89	18.64 (7.14)	18 (14-23)	(1-41)	16.84 (6.66)	17 (12.25-21)	(1-38)	20.08 (7.19)	20 (15-25)	(3-41)	.46***
ANGER	.84	15.35 (6.09)	15 (11-19)	(1-34)	14.76 (6.35)	14.5 (10-19)	(2-34)	15.83 (5.84)	15 (12-19)	(1-34)	.18
SADNESS	.80	18.23 (5.56)	18 (14-22)	(4-35)	16.81 (5.33)	17 (13-20)	(6-32)	19.38 (5.48)	20 (16-23)	(4-35)	.47***
Short version											
SEEKING	.75	12.69 (2.87)	13 (11-15)	(3-18)	12.66 (2.9)	13 (11-15)	(4-18)	12.72 (2.85)	13 (11-15)	(3-18)	.02
CARING	.68	12.07 (2.6)	12 (10-14)	(4-18)	11.29 (2.65)	11 (9-13)	(4-17)	12.7 (2.37)	13 (11-14)	(6-18)	.56***
PLAYFULNESS	.67	11.84 (2.68)	12 (10-14)	(5-18)	12.04 (2.59)	12 (10-14)	(5-18)	11.68 (2.74)	12 (10-14)	(5-18)	-0.13
FEAR	.82	7.39 (3.52)	7 (5-10)	(0-17)	6.46 (3.24)	6 (4-8)	(0-16)	8.15 (3.57)	8 (6-11)	(0-17)	.49***
ANGER	.81	6.97 (3.35)	7 (5-9)	(0-16)	6.53 (3.35)	6 (4-9)	(0-15)	7.33 (3.31)	7 (5-9)	(0-16)	.24***
SADNESS	.77	5.71 (2.99)	5 (4-8)	(0-16)	5.12 (2.89)	4.5 (3-7)	(0-14)	6.19 (2.98)	6 (4-8)	(0-16)	.36***

The table presents the descriptive statistics – mean and standard deviation (SD), median and interquartile range (IQR), minimum and maximum – for the total sample, by sex, and by ANPS version. The first column reports Cronbach's alpha (version for ordinal items). The last column indicates Hedges' g (effect size) for the differences between men and women. For each dimension, the range of possible scores is 0-42 for the ANPS-L and 0-18 for the ANPS-S. P-values refer to the t-test: *=p<.050; **=p<.010; *=p<.001

Table 3. Longitudinal measurement invariance models of the ANPS-L and ANPS-S

Measurement Invariance model (constraints)	Estimated parameters	Chi-square (DF)	CFI	RMSEA (90% CI)	ΔCFI	ΔRMSEA
Long version						
Configural (no equality constraints)	876	17652.275 (13654)	.825	.026 (.025-.027)		
Metric Invariance (loadings)	792	17704.444 (13738)	.827	.026 (.025-.027)	.002	.000
Scalar Invariance (loadings, thresholds)	632	17900.139 (13898)	.825	.026 (.025-.027)	-.002	.000
Residual Invariance (loadings, thresholds, residuals)	548	17947.036 (13982)	.827	.026 (.025-.027)	.000	.000
Complete Invariance (loadings, thresholds, residuals, means, var-cov)	521	17960.887 (14009)	.827	.026 (.025-.027)	.000	.000
Short version						
Configural (no equality constraints)	455	3649.369 (2316)	.897	.037 (.035-.039)		
Metric Invariance (loadings)	383	3774.053 (2388)	.893	.037 (.035-.039)	-.005	.000
Scalar Invariance (loadings, thresholds)	354	3857.161 (2417)	.889	.038 (.035-.040)	-.004	.001
Residual Invariance (loadings, thresholds, residuals)	318	3927.043 (2453)	.886	.038 (.036-.040)	-.003	.000
Complete Invariance (loadings, thresholds, residuals, means, var-cov)	291	4011.987 (2480)	.882	.038 (.036-.040)	-.004	.000

For each ANPS version, the table shows chi-square statistics and degrees of freedom (DF), the Comparative Fit Index (CFI), and the Root Mean Square Error Approximation (RMSEA) with 90% confidence intervals (90%CI) for each model. ΔCFI and ΔRMSEA refer to the difference between the model under consideration and the preceding (less constrained) model.

Table 4. Intraclass Correlation Coefficients of the ANPS-L and ANPS-S

		Long version				Short version			
	Group (N=341)	*Mean diff. (d)*	*Subject variance*	*Residual variance*	*ICC (95%CI)*	*Mean diff.*	*Subject variance*	*Residual variance*	*ICC (95%CI)*
SEEKING	**Global**	**.02**	**17.09**	**8.17**	**.68 (.60-.74)**	.03	**5.79**	**2.84**	**.67 (.60-.73)**
	Men	.04	17.13	9.47	.64 (.54-.74)	.07	5.93	3.13	.65 (.55-.75)
	Women	-.01	17.13	7.24	.70 (.61-.78)	-.01	5.72	2.64	.69 (.57-.77)
CARING	**Global**	**-.01**	**25.12**	**7.57**	**.77 (.73-.82)**	.06	**4.78**	**2.29**	**.68 (.61-.74)**
	Men	.04	21.06	7.67	.73 (.64-.80)	.12	4.22	2.57	.62 (.50-.74)
	Women	-.07	20.79	7.53	.73 (.67-.79)	.01	4.29	2.11	.67 (.58-.75)
PLAYFULNESS	**Global**	**.10**	**22.77**	**7.33**	**.76 (.70-.80)**	.13	**5.00**	**2.30**	**.69 (.61-.74)**
	Men	.10	23.61	7.83	.75 (.65-.82)	.11	4.78	2.46	.66 (.52-.76)
	Women	.10	22.10	7.00	.76 (.69-.81)	-.14	5.19	2.19	.70 (.61-.76)
FEAR	**Global**	**.09**	**39.06**	**11.35**	**.78 (.72-.81)**	.05	**8.49**	**3.26**	**.72 (.67-.77)**
	Men	.06	32.35	10.20	.76 (.67-.82)	-.02	6.39	3.35	.66 (.52-.75)
	Women	.10	39.09	12.22	.76 (.69-.82)	.09	8.60	3.18	.73 (.66-.79)
ANGER	**Global**	**.10**	**29.07**	**9.07**	**.76 (.71-.81)**	.08	**8.27**	**2.99**	**.74 (.68-.78)**
	Men	.07	30.38	9.87	.76 (.68-.82)	.10	8.23	3.14	.72 (.64-.79)
	Women	.11	23.61	7.83	.77 (.70-.82)	.06	8.24	2.86	.74 (.68-.80)
SADNESS	**Global**	**.10**	**22.00**	**10.69**	**.67 (.60-.73)**	.12	**5.42**	**3.82**	**.59 (.50-.67)**
	Men	0	19.99	8.69	.70 (.610-.765)	.05	4.91	3.30	.60 (.50-.68)
	Women	-.17	18.99	12.05	.61 (.494-.701)	.17	5.05	4.19	.55 (.39-.67)

The table shows Cohen's d (effect size) for the differences between T2 and T1 (none was significant according to the paired t-test), subject variance, residual variance, and Intraclass Correlation Coefficients (ICC) with 95% confidence intervals (95%CI) for the entire sample, by sex, and by ANPS version. ICC was used in the contingency form and calculated with the formula: *(Subject variance)/(Subject variance + Residual variance).*

Table 5. Models for measurement invariance across sex of the ANPS-L and ANPS-L

Measurement Invariance model (constraints)	Estimated parameters	Chi-square (DF)	CFI	RMSEA (90% CI)	ΔCFI	ΔRMSEA
Long version						
Configural (no equality constraints)	752	8647.072 (6720)	.812	.034 (.31-.36)		
Metric Invariance (loadings)	676	8695.018 (6796)	.814	.033 (.031-.035)	.002	-.001
Scalar Invariance (loadings, thresholds)	516	8898.938 (6956)	.810	.033 (.031-.035)	-.004	.000
Residual Invariance *(loadings, thresholds, residuals)*	*430*	*9085.062* *(7042)*	*.800*	*.034* *(.032-.036)*	*-.010*	*.001*
Residual partial Invariance * (loadings, thresholds, residuals)	431	9078.528 (7041)	.801	.034 (.032-.036)	-.009	.001
*Complete partial Invariance ** *(loadings, thresholds, residuals, means, var-cov)*	*398*	*9725.250* *(7128)*	*.746*	*.038* *(.036-.040)*	*-.055*	*.004*
Short version						
Configural (no equality constraints)	380	1528.256 (1092)	.919	.040 (.035-.044)		
Metric Invariance (loadings)	350	1558.960 (1122)	.919	.039 (.034-.044)	.000	-.001
Scalar Invariance *(loadings, thresholds)*	*286*	*1681.456* *(1186)*	*.908*	*.041* *(.036-.045)*	*-.010*	*.002*
Scalar partial Invariance ** (loadings, thresholds)	287	1664.557 (1185)	.911	.040 (.035-.044)	-.008	.001
Residual partial Invariance ** (loadings, thresholds, residuals)	251	1737.386 (1221)	.904	.041 (.036-.045)	-.007	.001
*Complete partial Invariance *** *(loadings, thresholds, residuals, means, var-cov)*	*218*	*1910.289* *(1254)*	*.878*	*.045* *(.041-.049)*	*.026*	*.004*

For each ANPS version, the table shows chi-square statistics and degrees of freedom (DF), the Comparative Fit Index (CFI), and the Root Mean Square Error Approximation (RMSEA) with 90% confidence intervals (90%CI) for each model. ΔCFI and ΔRMSEA refer to the difference between the model under consideration and the preceding (less constrained) model. Models in italics showed non-invariance.

* Free to vary in the second groups: residual variance of the item Anger 6 “When I am frustrated, I rarely become angry”

** Free to vary across groups: second threshold of the item Play 11 “I like all kinds of games including those with physical contact”