6-2019

Examining the cross-cultural validity of the positive affect and negative affect schedule between an Asian (Singaporean) sample and a Western (American) sample

Sean Teck Hao LEE
Singapore Management University, seanlee.2016@phdps.smu.edu.sg

Andree HARTANTO
Singapore Management University, andreeh@smu.edu.sg

Jose C. YONG

Brandon KOH

Angela K. Y. LEUNG
Singapore Management University, angelaleung@smu.edu.sg

Examining the cross-cultural validity of the positive affect and negative affect schedule between an Asian (Singaporean) sample and a Western (American) sample

Sean T. H. Lee, Andree Hartanto, Jose C. Yong, Brandon Koh and Angela K.-y. Leung

Abstract

The positive affect and negative affect schedule (PANAS) is a popular measure of positive (PA) and negative affectivity (NA). Developed and validated in Western contexts, the 20-item scale has been frequently administered on respondents from Asian countries with the assumption of cross-cultural measurement invariance. We examine this assumption via a rigorous multigroup confirmatory factor analysis, which allows us to assess between-group differences in both strength of scale item-to-latent factor relationship (metric invariance test) and mean of each scale item (scalar invariance test), on a large sample of 1,065 respondents recruited from Singapore (Asian sample) and the United States (Western sample). We found that two items assessing PA ("excited" and "proud") and three items assessing NA ("guilty," "hostile," and "ashamed") exhibited metric noninvariance whereas 11 of the remaining metric invariant items exhibited scalar noninvariance, suggesting that the PA and NA constructs differ from what the PANAS is expected to measure for Asian respondents. Our findings serve as a cautionary note to researchers who intend to administer the PANAS in future studies as well as to researchers interpreting the results of past studies involving respondents from Asian countries.

Keywords: cross-cultural comparison, measurement invariance, multigroup confirmatory factor analysis, negative affect, PANAS, positive affect.

Researchers have commonly employed the positive affect and negative affect schedule (PANAS)—a popular measure of positive and negative affectivity (i.e., PA and NA, respectively) consisting of 10 PA and 10 NA items (Watson, Clark, & Tellegen, 1988)—on Asian samples assuming that its items will be appraised similarly to their Western counterparts (Bagozzi, Wong, & Yi, 1999; Spencer-Rodgers, Peng, & Wang, 2010). However, the PANAS has only been rigorously validated with Western samples (e.g., Crawford & Henry, 2004; Merz et al., 2013). Established cultural differences in perceptions of emotions cast reasonable doubt over the comparability of responses on the PANAS between Asian and Western respondents. For instance, high-arousal positive emotions (e.g., excited) tend to be perceived less positively by Asian individuals with collectivistic tendencies relative to Western individuals with individualistic tendencies because such emotions are viewed as socially imposing; conversely, Asians tend to favour low-arousal positive emotions (e.g., calm) because such emotions are more conducive to social harmony (Tsai, 2007; Tsai, Miao, Seppala, Fung, & Yeung, 2007).

Self-conscious emotions are also appraised differently between cultures. Such emotions (e.g., shame) arise from reflections made on one's thoughts and behaviours in an interpersonal context (Leary, 2003; Tangney, 2005). Positive self-conscious emotions (e.g., proud) signal that one is doing well in life and is appraised positively by Westerners; however, Asians tend to perceive such emotions as hubris and appraise them less positively. In contrast to Westerners, Asians appraise negative self-conscious emotions (e.g., guilt) more favourably, as they perceive such emotions to signal the desire to admit wrongdoings and preserve relationships (Eid & Diener, 2001). Such differences in appraisal may then differentially impact the perceived frequency and extent to which these emotions are experienced (Boiger, De Deyne, & Mesquita, 2013; Boiger, Mesquita, Uchida, & Barrett, 2013). Despite such cultural variations, rigorous tests of measurement invariance for the PANAS have generally been limited to non-Asian samples (e.g., Gaudreau, Sanchez, & Blondin, 2006) whereas tests that have involved Asian participants are limited to mere attempts at validating shortened versions of the scale (e.g., Thompson, 2007).

To date, only one published study has attempted to assess the cross-cultural validity of the full scale using Asian samples (Weidong, Jing, & Schick, 2004). Although the authors found the PANAS to be noninvariant between Chinese and American respondents, their study has important limitations. First, they administered the PANAS in different languages between their comparison samples, potentially confounding their results with translational and semantic equivalence issues. Second, exploratory factor analysis (EFA) was used to test for measurement invariance instead of the more robust and established multigroup confirmatory factor analysis (multigroup CFA; Milfont & Fischer, 2015). Not only is EFA statistically inappropriate when there is a specific factor structure specified by theory, such as in the case of the PANAS (cf. Watson et al., 1988), it also does not allow for any direct between-group tests, and researchers have to rely solely on visual assessments of factor loadings to determine if the scale's items are equivalent between groups (Chen, Yang, & Morin, 2015; Henson & Roberts, 2006; Hurley et al., 1997). Multigroup CFA addresses these issues and allows us to directly assess between-group differences in both strength of scale itemto-latent factor relationship (metric invariance test) and mean of each scale item (scalar invariance test), providing us with a much more nuanced understanding of the scale's cross-cultural validity (Milfont & Fischer, 2015). Our study thus aims to conduct a robust test of measurement invariance using multigroup CFA on the full 20- item PANAS in its original form (i.e., in English) on large samples of English-proficient respondents from an Asian country and a Western country.

Method

We recruited our Asian sample from Singapore. The Singapore government emphasizes the use and mastery of English (its official language) while also taking steps to preserve Asian cultural heritage and core values amongst citizens (Gupta & Yeok, 1995; Pakir, 1991). Singaporean participants were then compared with U.S. participants representing the Western sample. There were no missing values, and no data exclusion had taken place. Our data and analyses syntax can be found at https://osf.io/dj97h/.

Participants

Asian sample. A total of 563 participants (171 males, 392 females, Mage = 21.7 years, SDage = 1.7 years) were recruited from a large Singaporean university. The Singaporean participants identified themselves as Chinese (n = 484; 86%), Malay (n = 13; 2.3%), Indian (n = 30; 5.3%), and "other" (n = 36; 6.4%). Western sample. A total of 502 participants (244 males, 258 females, Mage = 36.7 years, SDage = 12.2 years) were recruited from Amazon Mechanical Turk. We restricted participation to only individuals who were residing in the United States. The American participants identified themselves as Caucasian (n = 349; 69.5%), African American (n = 45; 8.9%), Hispanic (n = 22; 4.4%), Native American (n = 30; 6%), Asian (n = 42; 8.4%), and "other" (n = 14; 2.8%).

Measures

PA and NA. All 20 PA and NA items of the PANAS were administered, with response options ranging from 1 (very slightly or not at all) to 5 (extremely). The internal consistency of both PA items, (Cronbach's aAsian = .908, aWestern = .915, and aAll = .915), and NA items, (Cronbach's aAsian = .906, aWestern = .957, and aAll = .939), were found to be high.

Procedure

In accordance with Watson et al.'s (1988) original instructions, participants were asked to rate the extent to which they "generally feel this way" for each of the 20 PANAS items. Participants were also asked to provide demographic information (e.g., age).

Results and Analysis

Based on Milfont and Fischer's (2015) guidelines, we first conducted a CFA on each sample independently to ascertain that a two-factor structure (i.e., PA and NA each indicated by their respective 10 items) fits both data sets well. Following current measurement invariance studies (e.g., Jang et al., 2017), model fit was assessed based on three goodness-of-fit statistics: chi-square significance test, comparative fit index (CFI), and root mean square error of approximation (RMSEA). For CFI, values between .90 and .94 indicate acceptable fit whereas values of .95 or greater indicate good fit (Hu & Bentler, 1999). For RMSEA, values smaller than .08 indicate acceptable fit whereas values of .05 and smaller indicate good fit (Browne & Cudeck, 1992). We also report additional, relevant indices of standardized root mean square residual (SRMR; Schreiber, Nora, Stage, Barlow, & King, 2006), x (McNeish, An, & Hancock, 2018), c ^ and McDonald's noncentrality index (NCI; Cheung & Rensvold, 2002). Model fit was generally acceptable for both samples, v2 (169) = 733.924, p < .001, CFI = .926, RMSEA = .082, 90% CI [0.076, 0.088], SRMR = .060, x = .968, c ^ = .898, McDonald's NCI = .569 for the U.S. sample, and v2 (169) = 712.758, p < .001, CFI = .907, RMSEA = .076, 90% CI [0.070, 0.081], SRMR = .050, x = .952, c ^ = .911, McDonald's NCI = .616 for the Singapore sample.

Configural invariance was then assessed using multigroup CFA to ascertain that the two-factor structure fits equally well for both groups. All factor loadings and item intercepts were allowed to freely vary between both samples. Fit indices obtained suggest that the two-factor structure holds similarly well for both groups, v2 (338) = 1,446.682, p < .001, CFI = .917, RMSEA = .078, 90% CI [0.074, 0.083], SRMR = .055, xUS = .968, xSingapore = .952, c ^ = .829, McDonald's NCI = .594. Next, metric invariance was tested by constraining all factor loadings to be equivalent between groups while allowing the item intercepts to vary freely. Model fit was acceptable, v2 (356) = 1,545.053, p < .001, CFI = .911, RMSEA = .079, 90% CI [0.075, 0.083], SRMR = .065, xUS = .969, xSingapore = .952, c ^ = .818, McDonald's NCI = .572. However, a chi-square difference test revealed that the metric invariance model fit more poorly than did the configural model, Dv2 (18) = 98.371, p < .001. In other words, constraining factor loadings resulted in a significantly poorer fit as compared to when the factor loadings were free to vary, suggesting that some item-to-latent factor relationships were nonequivalent between the samples.

We then conducted a test of partial metric invariance by constraining each factor loading one at a time (instead of constraining all factor loadings simultaneously) to determine which specific items were exhibiting metric noninvariance between our samples (Byrne, Shavelson, & Muthen, 1989). Following previous measurement-invariance studies (e.g., Jang et al., 2017), the items with the least residual variance when loaded onto their respective latent factors were selected as reference items and constrained to be equal between groups. The items "interested" and "afraid" were selected as reference items for the latent PA and NA factors, respectively. Chi-square difference tests were then conducted, comparing model fit where only the factor loadings of "interested" and "afraid" were constrained with model fit where an additional item (i.e., one of the remaining 18 items) on top of "interested" and "afraid" was constrained. This was repeated until all items on the PANAS had been examined.

As summarized in Table 1, partial metric invariance testing revealed that the items “excited,” “proud,” “guilty,” “hostile,” and “ashamed” were noninvariant between the samples. The factor loadings (Table 2) indicated that the strength of relationship between the items “excited” and “proud” and their corresponding latent PA

Table 1
Partial Metric Invariance Test Results

Item	$\Delta\chi^2(1)$	p
Excited	7.639	.006
Strong	3.744	.053
Enthusiastic	2.228	.136
Proud	12.592	<.001
Alert	0.003	.956
Inspired	3.582	.058
Determined	2.083	.149
Attentive	0.027	.869
Active	0.485	.486
Distressed	2.696	.101
Upset	0.000	1.000
Guilty	11.020	.001
Scared	0.225	.635
Hostile	13.911	<.001
Irritable	0.139	.709
Ashamed	21.685	<.001
Nervous	1.865	.172
Jittery	0.083	.773

Table 2
Summary of Standardized Factor Loadings

	Standardized factor loadings			
Item	PA (U.S.)	PA (SG)	NA (U.S.)	NA (SG)
Interested	.648	.679		
Excited	.809	.751		
Strong	.739	.669		
Enthusiastic	.809	.801		
Proud	.779	.628		
Alert	.543	.607		
Inspired	.812	.738		
Determined	.688	.803		
Attentive	.604	.647		
Active	.749	.731		
Distressed			.811	.618
Upset			.840	.731
Guilty			.843	.670
Scared			.897	.803
Hostile			.793	.563
Irritable			.783	.617
Ashamed			.859	.673
Nervous			.781	.781
Jittery			.812	.697
Afraid			.894	.850

factor was significantly weaker for Singaporean respondents than for American respondents. This suggests that these items are less reflective of PA for Asians relative to Westerners. Similarly, the strength of the relationship between the items "guilty," "hostile," and "ashamed" and their corresponding latent NA factor was significantly weaker for Singaporean respondents relative to American respondents, indicating that these items are less reflective of NA for Asians than for Westerners.

We further tested whether partial scalar invariance could be established by further constraining the item intercepts for the items that had exhibited metric invariance (Milfont & Fischer, 2015; Steenkamp & Baumgartner, 1998). To do so, we constrained both the factor loadings and item intercepts of all items, except the metric noninvariant items of "excited," "proud," "guilty," "hostile," and "ashamed," to be equal between groups. Model fit was poor, v2 (371) = 1,847.635, p < .001, CFI = .890, RMSEA = .086, 90% CI [0.083, 0.090], SRMR = .078, xUS = .969, xSingapore = .952, c^ = .785, McDonald's NCI = .500. A chi-square difference test also revealed a significantly poorer fit compared to when only the factor loadings of the PANAS items (excluding the metric noninvariant items) were constrained, Dv2 (20) = 370.854, p < .001. This indicates that constraining the item intercepts resulted in a poorer fit as compared to when the item intercepts were free to vary, suggesting broad between-group nonequivalence in the origin or mean value of the items (Bialosiewicz, Murphy, & Berry, 2013).

To determine which tested items were exhibiting scalar noninvariance between our samples, we constrained each item's intercept one at a time and compared model fit to when none of the item intercepts were constrained. As summarized in Table 3, the items "interested," "strong," "alert," "inspired," "determined," "attentive," "active," "distressed," "upset," "irritable," and "jittery" were found to be noninvariant between groups,

Table 3
Partial Scalar Invariance Testing Results

Item	$\Delta\chi^2(1)$	p
Interested	17.799	<.001
Strong	32.003	<.001
Enthusiastic	0.270	.603
Alert	64.032	<.001
Inspired	31.905	<.001
Determined	30.566	<.001
Attentive	51.191	<.001
Active	27.813	<.001
Distressed	20.94	<.001
Upset	6.604	.010
Scared	3.115	.078
Irritable	12.006	.001
Nervous	1.891	.169
Jittery	5.751	.016
Afraid	2.488	.115

suggesting that the origin or mean value fundamentally differs between Singaporeans and Americans on these items. Hence, only the items "enthusiastic" (PA), "scared" (NA), "nervous" (NA), and "afraid" (NA) are truly invariant (exhibiting both metric and scalar invariance) between our Singaporean and American respondents.

Further analyses were conducted to examine whether age and gender accounted for the noninvariance of the 16 PANAS items (5 exhibiting metric noninvariance and 11 exhibiting scalar noninvariance). Multiple indicators– multiple causes (MIMIC) modelling was conducted whereby a potential covariate (e.g., age or gender) is specified as both a predictor of the latent PA and NA factors within a single model and a predictor of each of the PANAS items. According to Jang et al. (2017), when a predictor explains variance in a latent factor, which itself is reflected by its observed indicators (i.e., scale items), it indicates that the overall factor's score varies depending on the level of the predictor. Further, when a predictor explains variance in an observed indicator (i.e., a scale item), it indicates that responses on the corresponding scale item vary depending on the level of the predictor; thereby potentially explaining observed noninvariance in responses on the item in question.

We found that age was a significant predictor of both the latent constructs of PA, B = 3.293, SE = .669, p < .001, and NA, B = 3.642, SE = .453, p < .001. Further specifying age as a predictor for the 5 metric noninvariant items and the 11 scalar noninvariant items (separate models for each item) revealed age to be a significant predictor for four metric noninvariant items and six scalar noninvariant items: "excited," B = 5.559, SE = .493, p < .001; "proud," B = 1.989, SE = .441, p < .001; "alert," B = 2.682, SE = .409, p < .001; "determined," B = 0.098, SE = .015, p = .049; "attentive," B = 2.927, SE = .469, p < .001; "distressed," B = 1.771, SE = .465, p < .001; "guilty," B = 1.551, SE = .519, p = .003; "irritable," B = 0.947, SE = .440, p = .031; "ashamed," B = 2.728, SE = .535, p < .001; and "jittery," B = 1.556, SE = .471, p = .001. To examine if respondent's country (i.e., Singapore vs. the United States) explains the noninvariance of these implicated items beyond age, we created a dummy variable with "0" representing Singaporeans and "1" representing Americans and specified it as an additional predictor variable within the same models detailed earlier. After accounting for age, country remained a statistically significant predictor of both the latent constructs of PA, B = 0.196, SE = .030, p < .001, and NA, B = 0.061, SE = .020, p = .002. Likewise, country significantly explained variance beyond age in each of the four metric noninvariant items and six scalar noninvariant items implicated: "excited," B = 0.204, SE = .023, p < .001; "proud," B = 0.118, SE = .020, p < .001; "alert," B = 0.106, SE = .018, p < .001; "determined," B = 3.474, SE = .195, p < .001; "attentive," B = 0.087, SE = .021, p < .001; "distressed," B = 0.095, SE = .021, p < .001; "guilty," B = 0.099, SE = .023, p < .001; "irritable," B = 0.059, SE =0.020, p = .003; "ashamed," B = 0.165, SE = .024, p < .001; and "jittery," B = 0.053, SE = .021, p = .013

Gender, represented by a dummy variable coding male as "0" and female as "1," was found to be a significant predictor of PA, B = 0.138, SE = .030, p < .001, but not NA, B = 0.005, SE = .020, p = .818. Further specifying gender as a predictor of the PA items that exhibited noninvariance (i.e., the two metric noninvariant PA items and seven scalar noninvariant PA items) revealed gender to be a significant predictor for the two metric noninvariant PA items and three scalar noninvariant PA items: "interested," B = 0.069, SE = .026, p = .007; "excited," B = 0.055, SE = .023, p = .017; "proud," B = 0.074, SE = .019, p < .001; "determined," B = 0.069, SE = .023, p = .003; and "active," B = 0.053, SE = .022, p = .017.

To examine whether respondent's country explains the noninvariance of these implicated items beyond gender, we likewise specified the same country dummy variable as an additional predictor variable within the same models detailed earlier. After accounting for gender, country remained a statistically significant predictor of both the latent constructs of PA, B = 0.196, SE = .030, p < .001, and NA, B = 0.061, SE = .020, p = .002. However, while country accounted for variance of the two metric noninvariant PA items beyond gender: "excited," B = 0.204, SE = .023, p < .001, and "proud," B = 0.118, SE = .020, p < .001, country did not explain variance for three scalar noninvariant PA items beyond gender: "interested," B = 0.010, SE = .026, p = .706; "determined," B = 0.027, SE = .024, p = .261; and "active," B = 0.018, SE = .023, p = .425. Collectively, these results suggest that of 16 PANAS items exhibiting noninvariance between our Asian and Western respondents, 3 can potentially be explained by gender differences: "interested," "determined," and "active." For the remaining 13 PANAS items, respondent's country explains their noninvariance beyond age and gender differences between our samples. Our findings are summarized in Table 4.

Discussion

Our results highlight important cultural differences in the construal and experience of PANAS items. Specifically, while the two-factor model of PA and NA held up similarly well for both our U.S. and Singaporean samples (i.e., configural invariance), the factor loadings of the items "excited," "proud," "guilty," "ashamed," and

Table 4
Summary of Findings

Item	Metric invariance is observed	Scalar invariance is observed	Noninvariance possibly explained by gender differences
Interested (PA)	✓		✓
Excited (PA)			
Strong (PA)	✓		
Enthusiastic (PA)	✓	✓	
Proud (PA)			
Alert (PA)	✓		
Inspired (PA)	✓		
Determined (PA)	✓		✓
Attentive (PA)	✓		
Active (PA)	✓		✓
Distressed (NA)	✓		
Upset (NA)	✓		
Guilty (NA)			
Scared (NA)	✓	✓	
Hostile (NA)			
Irritable (NA)	✓		
Ashamed (NA)			
Nervous (NA)	✓	✓	
Jittery (NA)	✓		
Afraid (NA)	✓	✓	

"hostile" exhibited metric noninvariance between the samples. This means that the strength of the relationship between these items and their respective latent factors (i.e., PA or NA) differs significantly between our two groups of respondents, thereby calling into question the comparability of variances and covariances between respondents from Asian countries and those from Western countries on the aggregated PA and NA scores involving these items (e.g., regression analyses; Bialosiewicz et al., 2013; Byrne et al., 1989).

Examining each item's factor loading revealed that the strength of the relationship between the items "proud" and "excited" and their latent PA factor was weaker for Singaporean respondents relative to American respondents, suggesting that these items are less reflective of PA for Asian individuals relative to Western individuals. This corroborates past research on cross-cultural differences in emotion perceptions showing that both self-conscious, positively valenced emotions (i.e., proud) and high arousal, positively valenced emotions (i.e., excited) are construed less positively by Asian individuals than Western individuals (Eid & Diener, 2001; Tsai, 2007). Assuming that "excited" bears the minimum arousal level required to elicit noninvariance, it is surprising that "alert," a PA item that is associated with an even higher arousal level than "excited" (Other PA items are of lower arousal levels than "excited.") based on the circumplex model of affect, did not exhibit noninvariance in our study (Barrett & Russell, 1998). This is plausibly because “alert” is less positively valenced as compared to “excited,” as “alert” is considered as a close-to-neutral emotion in the valence dimension of the circumplex model (Barrett & Russell, 1998), although further empirical examination is needed to validate this postulation. Nonetheless, our results serve to caution researchers that certain PA items (i.e., “proud” and “excited”) on the PANAS may not be directly comparable between samples drawn from Asian countries and samples drawn from Western countries.

The strength of the relationship between the items “guilty,” “ashamed,” and “hostile” and their latent NA factor was also weaker for Singaporeans relative to Americans, suggesting that these items are less reflective of NA for Asians than for Westerners. This similarly corroborates previous research showing that self-conscious, negatively valenced emotions of guilt and shame are construed less negatively by Asians than by Westerners due to their perceived instrumentality in the maintenance of social harmony (Eid & Diener, 2001).

To explain the unexpected noninvariance of the item “hostile,” some researchers have argued that anger is a valuable source of social information because it signals the current status of interpersonal relations and allows for the modulation of behaviours to occur (Sell, Tooby, & Cosmides, 2009; van Kleef, 2009; van Kleef, Anastasopoulou, & Nijstad, 2010). In addition, research has shown that the display of anger and hostility may serve to reinforce one’s group identity when directed at outgroup members (Cikara, Botvinick, & Fiske, 2011; Schaafsma & Williams, 2012). Such social utility of this emotion may then render individuals from collectivistic cultures to appraise it less negatively than those from individualistic cultures, although, being an unanticipated finding, further empirical examination is needed to validate this. Nonetheless, these results serve to caution researchers that in addition to the PA items mentioned, certain NA items (i.e., “guilty,” “ashamed,” and “hostile”) on the PANAS may not be directly comparable between samples drawn from Asian countries and samples drawn from Western countries as well.

Further, we found that the PANAS item intercepts (i.e., origin or mean value) broadly differed between our samples. This suggests that baseline differences in the extent certain emotions are experienced on average between respondents from the two cultural contexts exist, calling into question the comparability of factor means between respondents from Asian countries and those from Western countries on aggregated PA and NA scores that involve the averaging of scores on the afflicted items (e.g., t tests; Bialosiewicz et al., 2013; Byrne et al., 1989; Lee, 2018). Among the 15 items that exhibited metric invariance, 11 of them exhibited scalar noninvariance between our samples, with “enthusiastic” (PA), “scared” (NA), “nervous” (NA), and “afraid” (NA) being the only items that exhibited both metric and scalar invariance. While scalar noninvariance observed for the items “interested,” “determined,” and “active” can potentially be explained by the samples’ gender compositional differences, respondent’s country still explained noninvariance in all other 13 items above and beyond age and gender differences. This suggests that while comparisons made on factor variances and covariances may still be acceptable should the five afflicted items exhibiting metric noninvariance be dropped from analyses, comparisons made on factor means between respondents with an Asian cultural background and those with a Western cultural background may not be substantiated at all (Bialosiewicz et al., 2013; Byrne et al., 1989). Further studies, though, are encouraged to validate and explicate these findings.

Overall, our findings are generally consistent with those of Weidong et al. (2004), in that the PANAS is noninvariant between respondents from an individualistic cultural context and those from a collectivistic cultural context. Adopting a more rigorous test of measurement invariance (i.e., multigroup CFA), we were able to identify significant differences in strength of certain item-to-latent factor relationships (i.e., metric noninvariant items) as well as baseline mean differences in the extent/magnitude that certain PANAS items are experienced between respondents from the two cultural contexts (i.e., scalar noninvariant items). This means that the latent constructs of PA and NA measured by aggregating these PANAS items broadly differ between these two groups of respondents, suggesting that any direct comparisons on PA and/or NA made between respondents from Asian countries and those from Western countries may be inherently confounded by cultural differences in emotional appraisal and experience, and that subgroup analyses may be necessary for researchers working with samples that include participants from both types of cultural background (Chan, 2011; Lee, 2018; Milfont & Fischer, 2015).

Recommendations and Conclusion

Our study presents the first empirical assessment of the full 20-item PANAS's cross-cultural validity on an Asian sample using multigroup CFA, thus elucidating the limitations of the current literature where the PANAS has been indiscriminately administered to Asian participants while assuming the scale measures the constructs of PA and NA similarly to that of Western participants. In addition, by assessing configural, metric, and scalar invariance, we were able to provide a nuanced understanding of the cross-cultural psychometric properties and validity of each PANAS item. This can guide future efforts aimed at improving the cross-cultural validity of the PANAS as well as reanalysing past studies that have administered the full PANAS on Asian samples.

Although the large sample sizes in the current study suggest that our findings are robust, future studies should aim to replicate our findings across a wider pool of Asian and Western countries to test for generalizability. Asian and Western participants matched by gender composition should also be recruited to determine whether the noninvariant items associated with gender effects would disappear. While we endeavoured to account for pertinent demographical differences between our two samples within our analyses, we acknowledge that other unaccounted demographical differences may exist (e.g., socioeconomic status) that may potentially impact our results. As these limitations stem primarily from resource constraints, we highly urge others with greater means to (a) replicate our findings with a large pool of sample from multiple countries and (b) match sampling method and demographics of respondents from all countries sampled. This would then allow us to draw a more robust conclusion regarding the cross-cultural validity of the PANAS between respondents with an Asian versus those with a Western cultural background.

That said, as our findings are generally consistent with past findings on cultural variations in the appraisal and perception of emotions, we can be fairly confident that our findings are both robust and valid. Overall, findings from our study serve as a cautionary note to researchers who intend to administer the PANAS in future studies as well as researchers interpreting the results of past studies that have used the PANAS on Asian samples and have made broad cross-cultural comparisons of PA and/or NA (e.g., Bagozzi et al., 1999; Spencer-Rodgers et al., 2010). Our findings also potentially limit the generalizability of past findings pertaining to PA and NA measured by the PANAS to the cultural context in which the study in question was conducted. We hope that findings from our study would serve to spur further empirical investigation on this issue and to guide future rectification efforts, which may include the identification of other culturally equivalent emotion terms to replace those that are deemed nonequivalent. Alternatively, researchers may also adopt other measures of PA and NA that do not directly reference any specific emotion, such as the affective slider (Betella & Verschure, 2016), when respondents from Asian countries are involved.