

The Meaning of Aggression Varies Across Culture: Testing the Measurement Invariance of  
the Refined Aggression Questionnaire in samples from Spain, the US, and Hong Kong

David Gallardo-Pujol<sup>1,2</sup>, Eva Penelo<sup>3</sup>, Cindy Sit<sup>4</sup>, Montsant Jornet-Gibert<sup>1</sup>, Carlos Suso<sup>5</sup>,  
Macià Buades-Rotger<sup>6</sup>, Alberto Maydeu-Olivares<sup>1,2,7</sup>, Antonio Andrés-Pueyo<sup>1,2</sup>, Fred B.  
Bryant<sup>8</sup>

<sup>1</sup>Universitat de Barcelona

<sup>2</sup>Institute for Neurosciences (UBNeuro)

<sup>3</sup>Laboratori d'Estadística Aplicada, Departament de Psicobiologia i Metodologia de les  
Ciències de la Salut, Universitat Autònoma de Barcelona, Barcelona, Spain.

<sup>4</sup>The Chinese University of Hong Kong

<sup>5</sup>Universitat Jaume I

<sup>6</sup>Universität zu Lübeck

<sup>7</sup>University of South Carolina

<sup>8</sup>Loyola University Chicago

ACCEPTED ON November, 13<sup>TH</sup> 2018

Correspondence should be addressed to: Dr. David Gallardo-Pujol, Dept. of Personality,  
Universitat de Barcelona, Edifici de Ponent - Campus Mundet, Pg. de la Vall d'Hebron 171,  
08035 Barcelona. Phone: +34 93 312 58 65. E-mail: david.gallardo@ub.edu

### **Acknowledgements**

This research was supported in part by the PSI2012-33601 grant, from the Spanish  
Ministry of Economy and Competitiveness, the Catalan Government Grant 2017SGR1237,  
and by the National Science Foundation Grant No. SES-1659936. The authors declare no  
conflict of interest.

## MEASUREMENT INVARIANCE OF THE AQ-R IN SPAIN, US AND HONG KONG

**Abstract**

Cultural differences in aggression are still poorly understood. The purpose of this article is to assess whether a tool for measuring aggression has the same meaning across cultures.

Analyzing samples from Spain (n=262), US (n=344) and Hong-Kong (n=645), we used confirmatory factor analysis to investigate measurement invariance of the refined version of the Aggression-Questionnaire (Bryant & Smith, 2001). The measurement of aggression was more equivalent between the Chinese and Spanish versions than between these two and the American version. Aggression does not show invariance at the culture level. Cultural variables such as affective autonomy or individualism may influence the meaning of aggression. Aggressive behavior models can be improved by incorporating cultural variables.

Keywords: *aggression, collectivism, affective autonomy, cultural differences, factor invariance*

## MEASUREMENT INVARIANCE OF THE AQ-R IN SPAIN, US AND HONG KONG

## The Meaning of Aggression Varies Across Culture: Testing the Measurement Invariance of the Refined Aggression Questionnaire in Spain, the US, and Hong Kong

Understanding factors that promote aggression is of crucial importance (Anderson & Bushman, 2002). Although aggression is inherent to human behavior (Anderson & Bushman, 2002; Berkowitz, 1993; Berkowitz & Harmon-Jones, 2004), little is known about how culture influences aggression (Fry, 1998).

Updating the Buss-Durkee Hostility Inventory (Buss & Durkee, 1957), Buss and Perry (1992) developed an empirically-based questionnaire, the Aggression Questionnaire (AQ), which became the gold standard in the field. They obtained a four-factor structure which has been replicated across samples (e.g. Ramirez, Andreu, & Fujihara, 2001; Williams et al., 1996) and languages (Andreu-Rodríguez, Peña-Fernández, & Graña-Gómez, 2002; Fossati, Maffei, Acquarini, & Di Ceglie, 2003; Maxwell, 2007; Nakano, 2001; von Collani & Werner, 2005).

In order to improve its structural stability, Bryant and Smith (2001) shortened the original AQ to 12 items (AQ-R). This version allows for efficient administration and maintains high standards of validity and reliability (Gallardo-Pujol et al., 2006). It has also been translated into Chinese (Maxwell, 2007) and Spanish (Gallardo-Pujol et al., 2006). Yet, its measurement invariance across culture remains unknown.

Measurement invariance or measurement equivalence consists of different levels (Kankaraš et al., 2010; Vijver & Leung, 1997). *Structural or configural invariance* exists when the given construct shows the same factor structure across different cultures. *Metric invariance* exists when factor loadings (which reflect the meaning of the construct) are equal across different cultures. Finally, *scalar invariance* exists when the intercepts of the

MEASUREMENT INVARIANCE OF THE AQ-R IN SPAIN, US AND HONG KONG  
indicators are the same across groups. This implies that mean differences across cultures may reflect actual mean differences in the latent constructs.

Many studies have explored the configural invariance of the AQ (Fossati et al., 2003; Gallardo-Pujol et al., 2006; Maxwell, 2007; Nakano, 2001; Vigil-Colet, Lorenzo-Seva, Codorniu-Raga, & Morales, 2005), confirming the same set of factors in all adaptations so far. Yet, its full measurement invariance (configural, metric, and scalar) across cultures has not been investigated. Establishing metric invariance is the first step in showing that cross-cultural differences in mean aggression scores reflect differences in aggression levels rather than unknown factors. Indeed, directly comparing mean scores (scalar invariance) without establishing metric invariance may produce distorted conclusions. Hence, the aim of this study was to evaluate measurement invariance across three different versions of the AQ-R: Spanish, American English, and Chinese (Hong Kong).

The reasons for choosing these three cultures are not trivial. Benet-Martínez (Aaker, Benet-Martínez, & Garolera, 2001; Benet-Martínez, 2007) proposed an approach for evaluating cultural differences based on a triangulation of three cultures that vary with respect to at least two explanatory constructs (Benet-Martínez, 2007). Hence, we selected samples from these three cultures because they vary on two socio-cultural dimensions (Schwartz & Bilsky, 1987). These dimensions describe preferences for one state of affairs over another that distinguish countries (Hofstede, 2001; Hofstede & McCrae, 2004). In this case, we evaluated individualism (the United States vs. Spain and Hong Kong) and affective autonomy (Hong Kong vs. Spain and the United States). *Individualism* (vs. collectivism), defined as the preference for a framework in which individuals are expected to take care of themselves (Hofstede, 2001; Hofstede & McCrae, 2004), has been linked to violence and aggression in Western societies (Menzer & Torney-Purta, 2012). Affective autonomy refers to the independent pursuit of affectively positive experiences (Schwartz & Bilsky, 1987); *high*

MEASUREMENT INVARIANCE OF THE AQ-R IN SPAIN, US AND HONG KONG  
*affective autonomy* is related to leading a pleasant, happy, and exciting life. Hence, *low affective autonomy* may be related to unhappiness, poor emotion regulation, frustration, and therefore proneness to exhibit aggressive behaviors (Matsumoto, Yoo, & Nakagawa, 2008).

The present analysis differs from earlier work in two ways: i) it is the first study to test the measurement invariance of the AQ-R across eastern and western cultures; and ii) it systematically selected three cultures that differ in terms of the possible explanatory or mediating variables responsible for observed structural differences.

## Method

### Participants and Procedure

The Spanish sample, taken from Gallardo-Pujol et al. (2006), consisted of 262 students from Catalonia (154 females, 99 males, and 9 who did not report gender). Mean age was 21.68 ( $SD = 2.84$ ). Further details are available in Gallardo-Pujol et al. (2006).

The US sample, taken from Bryant and Smith (2001), consisted of 344 American undergraduates (250 females and 94 males) at a private Midwestern metropolitan university. Mean age was 18.49 ( $SD = 1.26$ ). Further details are available in Bryant and Smith (2001).

The Hong Kong sample, taken from Maxwell (2007), consisted of 645 undergraduate Hong Kong Chinese students (372 females, 272 males, and 1 who did not report gender) at the University of Hong Kong. Mean age was 19.71 ( $SD = 1.26$ ). Further details are available in Maxwell (2007).

For all samples, participation was voluntary and anonymous, and all participants provided informed consent for the inclusion of their data. The analyses conducted in this study are secondary to already existing data. Secondary analyses involve reanalyzing data collected with different purposes, in order to pursue a new research question not addressed by the original study.

### Measures

## MEASUREMENT INVARIANCE OF THE AQ-R IN SPAIN, US AND HONG KONG

*Aggression Questionnaire-Refined* version (AQ-R; Bryant & Smith, 2001). This short self-report questionnaire consists of 12 Likert-type items rated on a 5-point<sup>1</sup> the scale ranging from 1 (*Never*) to 5 (*Always*). The AQ-R is organized in four scales of three items each: Physical Aggression (PA), Verbal Aggression (VA), Anger (ANG) and Hostility (HO). All versions showed good psychometric properties (Bryant & Smith, 2001; Gallardo-Pujol et al., 2006; Maxwell, 2007).

### Statistical Analysis

Multigroup confirmatory factor analysis was conducted using polychoric correlations with Diagonally Weighted Least Squares (WLSMV) with a mean and variance adjusted chi-square test as implemented in Mplus 7.2 (Muthén & Muthén, 2016). For model identification, factor loadings of the first item for each factor were freely estimated, but all factor variances were fixed at 1, to avoid the use of a marker item (Kim & Yoon, 2011). Factors were allowed to intercorrelate. Factorial invariance across the three samples was tested with the chi-square test (Asparouhov & Muthén, 2006) for nested models (Byrne, 2011; Vandenberg & Lance, 2000) estimated using mean and variance corrected statistics. This is the procedure DIFFTEST implemented in Mplus. We started with a configural model (Model 1), in which

---

<sup>1</sup> In order to compare all three questionnaires, we decided to recode values of “6” into values of “5” for the US sample, given that the frequencies of “6” responses were extremely low, median frequency = 4%, compared with the total sample. Converting the 6-point scale to a 5-point scale by recoding 6's as 5's produced item scores that were virtually identical ( $r_s \geq .988$ ) to those produced by subtracting 1 from 6-point-scale scores, multiplying the result by 0.8, and adding 1 to the product to obtain a 5-point scale. Additionally, Spain and Hong-Kong samples retained the original AQ 5-point rating scale that was modified in the US AQ-R. To make sure that recoding category data did not affect the results obtained, we repeated all measurement invariance analyses using the original coding (6 point scale for US and 5 point scale for China and Spain). The results obtained are consistent with those reported here, only partial measurement invariance holds, and for the same items and combinations of countries reported in this brief report.

MEASUREMENT INVARIANCE OF THE AQ-R IN SPAIN, US AND HONG KONG

all parameters were freely estimated across samples but the same theoretical model was specified across populations. Then, full metric invariance was tested (Model 2), by equating factor loadings across populations, and freeing factor variances in the second and third groups (which had been fixed at 1 in the first group for model identification, as in Ezpeleta & Penelo, 2015). The metric invariance model across the three populations was rejected. Then, we tested full metric invariance across two of the populations (Models 3 to 5). Then, we examined partially invariant models (Models 6 to 9) in which the parameters of one item were relaxed sequentially using a backward procedure (Kim & Yoon, 2011). Finally, should metric invariance was met; scalar invariance would have been explored. Goodness-of-fit was assessed using (Jackson, Gillaspay, & Purc-Stephenson, 2009):  $\chi^2$ , comparative fit index (CFI), Tucker and Lewis index (TLI), and root mean squared error of approximation (RMSEA) using conventional thresholds (Marsh, Hau, & Wen, 2004).

### Results

Table 1 reports descriptive statistics for each item and subscale, and internal consistency for each dimension in each of the three samples.

#### INSERT TABLE 1

Table 2 summarizes the results for the tests of measurement invariance across the three samples<sup>2</sup>. Full metric invariance did not hold across all three samples (Model 2) or between any of the three pairs of samples (Models 3 to 5). Partial metric invariance held

---

<sup>2</sup> Gender invariance was also tested within each country, given the asymmetry between males and females in terms of aggression. We found absolute gender invariance in the US ( $\chi^2=7.123$ ,  $df=8$ ,  $p\text{-value}=0.5234$ ) and Hong Kong ( $\chi^2=3.887$ ,  $df=8$ ,  $p\text{-value}=0.8672$ ). There was partial gender invariance (20% freed parameters) for the Spanish sample ( $\chi^2=6.938$ ,  $df=6$ ,  $p\text{-value}=0.3266$ ). The two items involved were one from VA scale (My friends say that I'm somewhat argumentative./Mis amigos/as dicen que soy discutidor/ra), and another one from the HOST scale (My friends say that I'm somewhat argumentative./Mis amigos/as dicen que soy discutidor/ra). In both cases, females had larger factor loadings than males.

MEASUREMENT INVARIANCE OF THE AQ-R IN SPAIN, US AND HONG KONG across pairs of samples as follows: 6 factor loadings equivalent for Spanish and US samples (Model 6), 8 factor loadings equivalent for Spanish and Hong Kong samples (Model 7), and 6 factor loadings equivalent for US and Hong Kong samples (Model 8). Finally, analysis of partial metric invariance across the three samples was conducted simultaneously. Partial metric invariance could not be rejected,  $\Delta\chi^2(6) = 12.3, p = .05$ . Fit statistics for the final, partially-invariant model (Model 9) were:  $\chi^2(157) = 554.7$ , CFI = .96, TLI = .94, and RMSEA = .080. Each model always included a multigroup approach, assessing all three groups, but just fixing parameters across two of the samples and freeing the third not involved (detailed results of sequential analyses are available upon request).

#### INSERT TABLE 2

Figure 1 shows standardized (unstandardized factor loadings are available upon request) factor loadings and factor correlations for the final partially-invariant model (Model 9). Equivalent factor loadings between samples were as follows: 5 items (2 for PA and 1 for each of the other factors) across Spanish and American samples, 5 items (all 3 for PA and 1 for VA and HO) across American and Hong Kong samples, and 7 items (2 for PA, VA and AN, and 1 for HO) across Spanish and Hong Kong samples. Of these, 4 items showed equivalent factor loadings across the three samples: 2 for PA, 1 for VA, and 1 for AN). In contrast, 2 items did not have equivalent factor loadings across any of the three samples: 1 for VA and 1 for AN, being items showing lower loadings in the Spanish sample. That HO was the only AQ factor with no equivalent loadings across all three samples suggests that culture influences the meaning of hostility more than the meaning of physical or verbal aggression or of anger—a conclusion consistent with cross-cultural research using the 29-item AQ (Vigil-Colet et al., 2005, p. 607).

#### INSERT FIGURE 1

### Discussion



## MEASUREMENT INVARIANCE OF THE AQ-R IN SPAIN, US AND HONG KONG

Our aim was to assess metric invariance across three versions of the AQ-R. Of the 12 AQ-R items, 7 (58.3%) were metric invariant for the Spanish and Chinese samples; whereas only 5 (41.6%) were invariant for the Spanish and Chinese samples, and 5 (41.6%) for the US and Chinese samples. This pattern of results suggests that aggression is closer in meaning between the Chinese and Spanish versions than between each of these two versions and the American version.

One potential explanation for discrepancies is the use of an imposed-etic approach. This approach refers to the generalized practice of translating and adapting items originally adapted within one culture to another one, in contrast to an emic approach, that relies on items originally developed from within that culture (Berry, 1980; Berry, Poortinga, Breugelmans, Chasiotis, & Sam, 2011). Although imposed-etic instruments allow for quick comparisons across cultures, measurement was not metric equivalent in all three countries, suggesting that the meaning of aggression differs across culture. However, this does not explain the similarities between the two adaptations from English into Spanish and Chinese. These cross-cultural similarities may be attributed to certain values present in each of these societies (Schwartz, 1992). In particular, the similarity between Spain and Hong Kong with respect to PA and VA subscales may be explained by the similarity between both cultures with respect to individualism (Menzer & Torney-Purta, 2012). Collectivistic societies report fewer episodes of violence at schools (Menzer & Torney-Purta, 2012). Spanish and US adaptations are closer when considering AN and HO subscales. Thus, it might be reasonable to think that these societies conceive and promote both aspects of aggression in a similar way, given that Spain and the US show similar levels of affective autonomy compared to Hong Kong society (Aaker et al., 2001, p. 494). However, the variables studied here cannot explain the high degree of variation that remains across all three cultures with respect to the self-reported manifestations of aggression.

## MEASUREMENT INVARIANCE OF THE AQ-R IN SPAIN, US AND HONG KONG

Future research should include comparisons among cultures differing on other cultural dimensions (Schwartz, 1992). Such research would complement current aggression models (e.g., the General Aggression Model) which do not go beyond proximal causes of aggression (Anderson & Bushman, 2002). Moreover, because contemporary models of aggression are culturally-centered within the perspective of Western societies (Henrich, Heine, & Norenzayan, 2010), it is important to develop cross-cultural models. Additionally, an important avenue of research could be using IRT analyses to study cross-cultural differences in AQ-R (and other measures) with respect to differential item functioning or differential test functioning. It is likely that in the future it would enable fine-grained comparisons (e.g. Hambrick et al., 2010). The present work is not exempt from limitations that should be addressed in further studies. We found that mean age was different in all three samples. This could actually be affecting the composition of the sample and thus hampering the robustness of our findings. However, there is evidence that by the age of our subjects, aggression has already peaked in late adolescence and is actually slowly steadily declining at similar levels (Liu, Lewis, & Evans, 2012; Moffitt, 1993). With respect to gender, we conducted separate analyses to explore gender invariance (see footnote 2) within each country. We only found partial invariance in Spain, but at the threshold for accepting it in practical applications (Dimitrov, 2010), as it is intended this questionnaire (Gallardo-Pujol et al., 2006).

All in all, our results have shown that 1) metric invariance should be tested before proceeding to direct comparisons of national/cultural mean levels of aggression, and 2) certain cultural variables, such as individualism and affective autonomy, may influence the meaning of aggression across culture (Schwartz, 1994). As has typically been the case in previous comparative cross-cultural research on the AQ, the present study did not assess criterion measures as correlates of AQ-R subscales across multiple countries. However, because such criterion measures are crucial for establishing cross-cultural construct validity,

MEASUREMENT INVARIANCE OF THE AQ-R IN SPAIN, US AND HONG KONG

future international work on the AQ-R should include criterion measures. Our results suggest that this future research should be careful to address potential cross-cultural differences in factor structure, which could otherwise produce misleading evidence about the generalizability of construct validity across culture.

### References

- Aaker, J. L., Benet-Martínez, V., & Garolera, J. (2001). Consumption symbols as carriers of culture: A study of Japanese and Spanish brand personality constructs. *Journal of Personality and Social Psychology, 81*(3), 492–508.
- Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology, 53*, 27–51.
- Andreu-Rodríguez, J. M., Peña-Fernández, M. E., & Graña-Gómez, J. L. (2002). Adaptación psicométrica de la versión española del Cuestionario de Agresión. *Psicothema, 14*(2), 476–482.
- Asparouhov, T., & Muthén, B. (2006). Robust Chi Square Difference Testing with Mean and Variance Adjusted Test Statistics. *Mplus Web Notes: No 10*.
- Benet-Martínez, V. (2007). Cross-cultural personality research: Conceptual and methodological issues. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 170–189). New York: Guilford Press.
- Berkowitz, L. (1993). *Aggression: Its causes, consequences and control*. New York: McGraw-Hill.
- Berkowitz, L., & Harmon-Jones, E. (2004). Toward an understanding of the determinants of anger. *Emotion, 4*(2), 107–130.
- Berry, J. W. (1980). Introduction to methodology. In H. Triandis & J. W. Berry (Eds.), *Handbook of cross-cultural psychology* (Vol. 2, pp. 1–28). Boston: Allyn & Bacon.
- Berry, John W., Poortinga, Y. H., Breugelmans, S. M., Chasiotis, A., & Sam, D. L. (2011). *Cross-cultural psychology: Research and applications* (p. 650). Cambridge University Press.

## MEASUREMENT INVARIANCE OF THE AQ-R IN SPAIN, US AND HONG KONG

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity.

*Psychological Review*, *111*(4), 1061–1071.

Bryant, F. B., & Smith, B. D. (2001). Refining the architecture of aggression: A measurement model for the Buss-Perry Aggression Questionnaire. *Journal of Research in Personality*, *35*, 138–167.

Buss, A. H., & Durkee, A. (1957). An inventory for assessing different kinds of hostility.

*Journal of Consulting Psychology*, *21*, 343–349.

Buss, A. H., & Perry, M. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology*, *63*(3), 452–459.

Byrne, B. M. (2011). *Structural equation modeling with Mplus: Basic concepts, applications, and programming (Multivariate Applications Series)* (p. 430). New York: Taylor & Francis.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct

Validation. *Measurement and Evaluation in Counseling and Development*, *43*(2), 121–149. <http://doi.org/10.1177/0748175610373459>

Ezpeleta, L., & Penelo, E. (2015). Measurement Invariance of Oppositional Defiant Disorder

Dimensions in 3-Year-Old Preschoolers. *European Journal of Psychological Assessment*, *31*(1), 45–53. <https://doi.org/10.1027/1015-5759/a000205>

Fossati, A., Maffei, C., Acquarini, E., & Di Ceglie, A. (2003). Multigroup confirmatory component and factor analyses of the Italian version of the Aggression Questionnaire.

*European Journal of Psychological Assessment*, *19*(1), 54–65.

Fry, D. P. (1998). Anthropological perspectives on aggression: Sex differences and cultural variation. *Aggressive Behavior*, *24*(2), 81–95. doi:10.1002/(SICI)1098-

2337(1998)24:2<81::AID-AB1>3.0.CO;2-V

## MEASUREMENT INVARIANCE OF THE AQ-R IN SPAIN, US AND HONG KONG

Gallardo-Pujol, D., Kramp, U., García-Forero, C., Perez-Ramirez, M., Andres-Pueyo, A.,

Kramp, U., et al. (2006). Assessing aggressiveness quickly and efficiently: The Spanish adaptation of Aggression Questionnaire-Refined version. *European Psychiatry*, *21*(7), 487–494.

Hambrick, J. P., Rodebaugh, T. L., Balsis, S., Woods, C. M., Mendez, J. L., & Heimberg, R.

G. (2010). Cross-Ethnic Measurement Equivalence of Measures of Depression, Social Anxiety, and Worry. *Assessment*, *17*(2), 155–171.

<http://doi.org/10.1177/1073191109350158>

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?

*Behavioral and Brain Sciences*, *33*(2-3), 61–83. doi:doi:10.1017/S0140525X0999152X

Hinds, P.S., Vogel, R.J., Clarke-Steffen, L. (1997) 'The possibilities and pitfalls of doing a secondary analysis of a qualitative data set', *Qualitative Health Research*, vol. 7(3): 408-24.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations* (p. 616). SAGE.

Hofstede, G., & McCrae, R.R. (2004). Personality and Culture Revisited: Linking Traits and Dimensions of Culture. *Cross-Cultural Research*, *38*(1); 52-88. doi:

[10.1177/1069397103259443](http://doi.org/10.1177/1069397103259443)

Jackson, D. L., Gillaspay, J. A., & Purc-Stephenson, R. (2009). Reporting practices in

confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*(1), 6–23. doi:10.1037/a0014694

Kankaras, M., Vermunt, J. K., & Moors, G. (2011). Measurement Equivalence of Ordinal Items: A Comparison of Factor Analytic, Item Response Theory, and Latent Class

Approaches. *Sociological Methods & Research*, *40*(2), 279–310.

<http://doi.org/10.1177/0049124111405301>

## MEASUREMENT INVARIANCE OF THE AQ-R IN SPAIN, US AND HONG KONG

- Kim, S.-H., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling, 18*, 212-228.  
doi:10.1080/10705511.2011.557337
- Liu, J., Lewis, G., & Evans, L. (2012). Understanding aggressive behaviour across the lifespan. *Journal of psychiatric and mental health nursing, 20*(2), 156-68.
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320–341. doi:10.1207/s15328007sem1103\_2
- Matsumoto, D., Yoo, S. H., & Nakagawa, S. (2008). Culture, emotion regulation, and adjustment. *Journal of Personality and Social Psychology, 94*(6), 925–937.  
doi:10.1037/0022-3514.94.6.925
- Maxwell, J. P. (2007). Development and preliminary validation of a Chinese version of the Buss-Perry Aggression Questionnaire in a population of Hong Kong Chinese. *Journal of Personality Assessment, 88*(3), 284–94. doi:10.1080/00223890701317004
- Menzer, M. M., & Torney-Purta, J. (2012). Individualism and socioeconomic diversity at school as related to perceptions of the frequency of peer aggression in fifteen countries. *Journal of adolescence. doi:10.1016/j.adolescence.2012.04.013*
- Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100*(4), 674-701. doi:10.1037/0033-295X.100.4.674
- Muthén, L. K., & Muthén, B. O. (2016). *Mplus. Version 7.2* [Computer program]. Los Angeles, CA: Muthén & Muthén.
- Nakano, K. (2001). Psychometric evaluation of the Japanese adaptation of the Aggression Questionnaire. *Behaviour research and therapy, 39*(7), 853–8.

## MEASUREMENT INVARIANCE OF THE AQ-R IN SPAIN, US AND HONG KONG

- Ramirez, J. M., Andreu, J. M., & Fujihara, T. (2001). Cultural and sex differences in aggression: A comparison between Japanese and Spanish students using two different inventories. *Aggressive Behavior*, 27(4), 313–322. doi:10.1002/ab.1014
- Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1–65). San Diego, CA: Academic Press.
- Schwartz, S. H., & Bilsky, W. (1987). Toward a universal psychological structure of human values. *Journal of Personality and Social Psychology*, 53(3), 550–562.  
<http://doi.org/10.1037/0022-3514.53.3.550>
- 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), 4–70. doi:10.1177/109442810031002
- Vigil-Colet, A., Lorenzo-Seva, U., Codorniu-Raga, M. J., & Morales, F. (2005). Factor structure of the Buss-Perry Aggression Questionnaire in different samples and languages. *Aggressive Behavior*, 31(6), 601–608.
- Vijver, F. J. R. van de, & Leung, K. (1997). *Methods and data analysis for cross-cultural research (Cross Cultural Psychology)*. Thousand Oaks, CA: Sage Publications.
- von Collani, G., & Werner, R. (2005). Self-related and motivational constructs as determinants of aggression. *Personality and Individual Differences*, 38(7), 1631–1643.  
doi:10.1016/j.paid.2004.09.027
- Williams, T. Y., Boyd, J. C., Cascardi, M. A., & Poythress, N. (1996). Factor structure and convergent validity of the Aggression Questionnaire in an offender population. *Psychological Assessment*, 8(4), 398–403.



## MEASUREMENT INVARIANCE OF THE AQ-R IN SPAIN, US AND HONG KONG

Table 1

*Item and scale descriptives for Spanish, American, and Hong Kong samples*

Scale Item	Mean (SD)			Internal consistency (mean inter-item correlation)			
	Spain (n = 262)	US (n = 344)	Hong Kong (n = 645)	Spain	US	Hong Kong	Total
Physical Aggression	4.74 (1.94)	5.72 (3.31)	4.21 (1.71)	.76 (.51)	.78 (.55)	.68 (.43)	.76 (.51)
03	1.46 (0.72)	1.76 (1.16)	1.53 (0.81)				
08	1.58 (0.82)	2.06 (1.42)	1.29 (0.64)				
11	1.69 (0.82)	1.89 (1.35)	1.40 (0.73)				
Verbal Aggression	8.42 (1.91)	8.52 (3.53)	6.56 (2.22)	.52 (.28)	.81 (.59)	.67 (.40)	.74 (.49)
01	2.94 (0.60)	3.21 (1.18)	2.27 (0.86)				
06	2.55 (1.07)	2.71 (1.51)	2.30 (0.96)				
09	2.93 (0.93)	2.60 (1.44)	1.99 (1.03)				
Anger	7.21 (2.08)	7.25 (3.50)	6.42 (2.29)	.60 (.34)	.76 (.52)	.71 (.46)	.67 (.41)
05	2.38 (0.97)	2.40 (1.44)	2.48 (1.04)				
07	2.98 (0.99)	2.85 (1.50)	1.79 (0.87)				
12	1.85 (0.83)	2.00 (1.30)	2.16 (0.95)				
Hostility	6.37 (2.13)	7.96 (3.21)	6.56 (2.17)	.74 (.49)	.73 (.48)	.55 (.30)	.67 (.40)
02	1.98 (0.83)	2.70 (1.21)	1.88 (0.91)				
04	1.95 (0.89)	2.69 (1.44)	2.11 (0.98)				
10	2.43 (0.90)	2.57 (1.33)	2.56 (1.09)				

## MEASUREMENT INVARIANCE OF THE AQ-R IN SPAIN, US AND HONG KONG

Table 2

*Multigroup confirmatory factor analysis: Measurement equivalence across Spanish, American, and Hong Kong samples*

Model	Goodness-of-fit indices				Comparison*		
	$\chi^2$ (df)	CFI	TLI	RMSEA	Models	$\Delta \chi^2$ ( $\Delta df$ )	$p$
1. Same configuration across all three groups	570.733 (144)	.952	.934	.084			
2. Total metric invariance ( $\lambda$ ) across all three groups	700.681 (160)	.939	.925	.090	2 vs. 1	129.987 (16)	<.005
3. Total metric invariance ( $\lambda$ ) Spanish-American	593.367 (152)	.950	.935	.083	3 vs. 1	41.302 (8)	<.005
4. Total metric invariance ( $\lambda$ ) Spanish-Hong Kong	622.746 (152)	.947	.931	.086	4 vs. 1	59.748 (8)	<.005
5. Total metric invariance ( $\lambda$ ) American-Hong Kong	669.677 (152)	.942	.924	.090	5 vs. 1	80.320 (8)	<.005
6. Partial metric invariance ( $\lambda$ ) Spanish-American	560.125 (149)	.954	.939	.081	6 vs. 1	8.911 (5)	.113
7. Partial metric invariance ( $\lambda$ ) Spanish-Hong Kong	551.105 (152)	.955	.942	.079	7 vs. 1	15.051 (8)	.058
8. Partial metric invariance ( $\lambda$ ) American-Hong Kong	554.436 (149)	.954	.939	.081	8 vs. 1	10.820 (5)	.055
9. Partial metric invariance ( $\lambda$ ) across all three groups	554.672 (150)	.955	.940	.080	9 vs. 1	12.246 (6)	.057

\*Based on difference chi-square test for mean and variance adjusted chi-squares (Asparouhov & Muthén, 2006) .

MEASUREMENT INVARIANCE OF THE AQ-R IN SPAIN, US AND HONG KONG

Figure 1. Standardized factor loadings and factor correlations for Spanish (left), US (center) and Hong Kong (right) samples after testing for metric invariance of unstandardized factor loadings across the three countries. Error variances are omitted. Loadings in bold font were equivalent across groups. Square brackets indicate original item numeration.

Note. PA = Physical Aggression; VA = Verbal Aggression; ANG = Anger; HOS = Hostility.

