PSICOLOGIA

Revista da Associação Portuguesa de Psicologia
ISSN 2183-2471

Revista PSICOLOGIA, Vol. 37 (1), 2023, 53-63. doi: https://doi.org/10.17575/psicologia.1792

Latent mean differences between men and women: The case of the Preference for the Intuition and Deliberation Scale

Alejandro César Cosentino [1,2] & Susana Celeste Azzollini [2,3,4]

[1] *Colegio Militar de la Nacion (CMN)*
[2] *Centro de Investigaciones Humanas y Sociales para la Defensa (CISOHDEF), Facultad del Ejercito (FE), Universidad de la Defensa Nacional (UNDEF)*
[3] *Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)*
[4] *Universidad de Buenos Aires (UBA)*

Abstract: Intuition and deliberation are two modes of thinking for decision making. The objective of this research was to compare latent means between men and women's preference for intuition and deliberation. However, empirical studies on the Preference for Intuition and Deliberation scale (PID) measurement invariance were not available. The results of our study showed the original PID-based model did not show a good fit to the data. Nevertheless, a revised PID-based model showed strong and strict measurement invariance. As a result, latent mean comparison indicated that women showed more preference for intuition and less for deliberation than men.

Keywords: *Decision making; human sex differences; structural equation modeling; effect size.*

The difference between men and women in psychological human attributes is a topic that continues to be studied in psychology. Research based on 216 meta-analyses and large-sample studies, for example, described differences between men and women with magnitudes ranging from small to large in several psychological attributes. Examples of these include object location memory, mental rotation, the drive to analyze rule-based systems, negative emotions, anxiety disorders, impulsivity, social interests and abilities, escalated aggression, sex drive, interest in casual sex, socio-sexuality, and sexual violence (Archer, 2019).

One of the most interesting aspects of the descriptive study of the differences in psychological characteristics between men and women is that it drives the exploration of the mechanisms or processes that cause them. Consistently, these differences are explained by social role researchers as a consequence of the differences in the processes of socialization and power, while the theorists' explanations of the evolutionary perspective assume that differences between men and women may involve other causal factors, such as some kind of prenatal hormone (Schmitt, 2015).

A t-test or a variant of it on the observed total scores obtained from a measuring instrument containing several items is a procedure usually used in the literature to describe differences in psychological variables between groups (e.g., Pachi et al., 2022). So, if the result shows a statistically significant difference, the researchers usually support the objective existence of differences between groups. However, such a procedure ignores a series of problems inherent in psychological measurement scales, such as the effect of their measurement errors. In consequence, the result of the contrast would be described as "comparing apples and spark plugs" (Vandenberg & Lance, 2000, p. 9).

To prevent misunderstandings in the contrast of psychological variables across groups that are measured with the usual instruments in psychology, such as inventories or questionnaires, it is necessary to establish a set of characteristics of these measurement instruments. The structural equation modeling is used to determine different aspects of the measurement invariance across groups such as men and women. The measurement invariance study includes determining the configural invariance, the factor loading invariance (weak invariance), the item threshold invariance (strong invariance), the measurement error variances invariance (strict invariance), and the factor means invariance. Using this perspective, measurement invariance across men and women was studied for psychological variables, such as coronavirus anxiety (Caycho-Rodríguez et al., 2022), work gratitude (Youssef-Morgan et al., 2022), sense of coherence (Grevenstein & Bluemke, 2022), self-determination theory-related dimensions (Abós et al., 2021), and domains of creativity (Miroshnik et al., 2022).

[1] Correspondence address: Alejandro César Cosentino, Yapeyú 92 5 10, C1202ACB, Ciudad Autónoma de Buenos Aires, República Argentina. E-mail: acosentino@outlook.com

In this research, we show a sequence of studies to illustrate the conditions to determine differences between men and women for the decision-making preference. Through the Preference for Intuition and Deliberation Scale measurement instrument adaptation case, we want to show that variable comparison between men and women is only accurate after we have performed some modifications to the original characteristics of an instrument. Finally, after a multistep procedure, we can precisely compare latent variable scores and scores derived directly from the adapted instrument.

How to find differences between men and women: The preference for intuition and deliberation scale adaptation case

According to Betsch and Iannello (2010), there are a number of measurement instruments to assess individual inclination (tendency) to decide based on intuition or on deliberation, such as the Cognitive Style Indicator (CoSI; Cools & Van den Broeck, 2007), the General Decision Making Style (GDMS, (Scott & Bruce, 1995), the Rational Experiential Inventory (REI, Pacini & Epstein, 1999) and the Preference for Intuition and Deliberation scale (PID; Betsch, 2004). PID is, particularly, a measure of the individual inclination to the basic modes of decision making (Betsch & Kunz, 2008) based on intuition or on deliberation. PID has been included in numerous empirical studies (e.g., Colombo et al., 2013; Dijkstra et al., 2017; Stevenson & Hicks, 2016) and in meta-analyses (Phillips et al., 2016; Wang et al., 2015).

Specifically, the research has found that women score themselves higher in the preference for intuition than men (Betsch, 2004; Mikušková et al., 2015), although no differences were found in preference for deliberation between men and women (Laborde et al., 2010). However, no studies on PID measurement invariance are available, and neither are there studies comparing latent means of preference for intuition and deliberation. Only by determining the existence of strict measurement invariance across men and women for PID can a reasonable comparison of the observed scores be justified. In addition, the determination of strong measurement invariance across men and women may in turn contribute to determining the existence of difference between groups in the factors, beyond the effect of measurement errors in the scores of PID subscales. Consequently, we have studied the measurement invariance and latent means comparison across men and women. Meeting this objective required adapting PID to Argentinian Spanish and examining its psychometric characteristics. Accordingly, three studies were done to achieve these goals.

STUDY 1

This study aimed to provide Argentinian Spanish items so as to develop a culturally adapted PID. This scale (Betsch, 2004) is used for measuring the individual inclination to basic modes of decision making strategies based on affect (intuition) and cognition (deliberation; Betsch & Iannello, 2010; Betsch & Kunz, 2008). PID presents two subscales: PID-I measures preference for intuition, and PID-D measures preference for deliberation. PID was originally published in German (Betsch, 2004), and versions in several languages, such as Italian (Iannello, 2008), Dutch, and English (Betsch, 2008) have also been published. No information is available on how these translations were made, nor has a Spanish version of PID been published before. The psychometric characteristics of PID have been studied by a number of researchers (Betsch, 2004; Betsch, 2008; Iannello, 2008; Laborde et al., 2010; Mikušková et al., 2015; Monacis et al., 2016; Richetin et al., 2007; Schunk & Betsch, 2006; Witteman et al., 2009). These studies show that the internal consistency for PID-I ranges from .62 to .87, and for PID-D from .75 to .84. In addition, the scores of the subscales correlate with each other from -.36 to .29. Studies with PID model have shown that correlated PID factors imply a better fit to data than a non-correlated factor model (Betsch, 2004; Iannello, 2008).

Exploratory factor analyses (EFAs) have been performed on PID scores. On the one hand, the study showing the development of PID used principal axis factor method with orthogonal Varimax rotation with Kaiser Normalization (Betsch, 2004). On the other, a validation study of the English PID version used Principal Component Analysis method with Oblimin rotation. Both studies used the eigenvalue > 1 criterion and, by means of the graphic method of the scree test, 2 factors were retained. In these studies, both factors explained 31.8% and 34.7% of the variance, respectively.

In the PID development study, the loading pattern structure obtained through EFA showed that less than half of the items present cross-loadings very close to 0 (Betsch, 2004). For the items that load more on the PID-I factor (loading range from .35 to .68), their cross-loadings on PID-D range from .06 to | .31 |, while those that load on the PID-D factor (loading range from .40 to .75) have cross-loadings ranging from | .03 | to .28. This may be explained by the fact that, on the one hand, various items correlate with both factors, and on the other, for most items, the item-to-factor loading on a specific factor could be classified as less than good (loadings < .55), according to the Comrey and Lee (1992) guidelines.

Therefore, we assume that PID does not present a simple or approximately simple structure, but at least an approximately complex structure (Schmitt & Sass, 2011).

The validation study of the PID English version suggested that some items could be considered problematic (Richetin et al., 2007). In particular, item 5 corresponding to the PID-I subscale (*I do not like situations that require me to rely on my intuition*, reversed scoring) presented a loading of .15 and a cross-loading of .12(in the original study, item 5 had a loading of .51 and a cross-loading of -.22), while item 6 corresponding to subscale PID-D (*I think about myself*) yielded a loading of .23 and a cross-loading of .09 (in the original study, item 6 had a loading of .52 and a cross-loading of .28). In contrast, a confirmatory factor analysis (CFA) in the study of the Italian PID version showed that item 5 presented a standardized loading factor of .35 with respect to PID-I, and that item 6 presented a standardized loading factor of .17 with respect to PID-D.

It should be noted that item 2 of PID-I (*I listen carefully to my deepest feelings*) presented an even more complex situation in the validation study of the PID English version. This item had a loading of .23 and, surprisingly, a cross-loading of .53, which implies that, contradictorily, it loaded more on the other factor than on the one that it would hypothetically correspond to (in the original study, item 2 had a loading of .53 and a cross-loading of .27; Betsch, 2004).

Given the type of structure that PID presented and the findings of suboptimal items, we assumed that the results of an EFA from the Argentinean Spanish PID version could have some weaknesses.

Method

Participants. A convenience sample of general population consisted of 240 participants (120 women) with a mean age of 39.2 years (*SD* = 14.7, age range from 17 [one participant] to 89). Respondents were volunteers that did not receive any compensation for their participation.

Materials and Procedure

Preference for Intuition and Deliberation scale (PID). This measurement instrument was developed by Betsch (2004) to assess the preferences for intuitive and for deliberative decision making. The PID consists of 9 items to measure preference for intuition (subscale PID-I) and 9 items to measure preference for deliberation (subscale PID-D). The items consist of statements, and the participants indicate their agreement using a 5-point scale ranging from 1 (I very much disagree) to 5 (I very much agree).

Procedures. To translate the PID version into Argentinian Spanish, the front-and back-translation technique was used. A general population sample was used to analyze the psychometric properties of the Argentinian Spanish PID version and to explore its two-factor structure by means of an exploratory factor analysis. The data were analyzed following the classical test theory approach. The FACTOR software was used to determine the number of factors to be extracted (Version 9.2; Lorenzo-Seva & Ferrando, 2013). A parallel analysis for determining the number of factors was conducted. It involved a minimum rank factor analysis (PA-MRFA) with a 95% quantile criterion threshold. This threshold was based on a model of common factors, on a polychoric correlations matrix –as responses to items consisted of 5 ordered categories and did not present a normal distribution. The CEFA software was used to calculate rotations (Version 3.04; Tateneni et al., 2009). Oblique rotations were selected because models with correlated PID factors result in a better fit to data. The MVN package (Version 4.0; Korkmaz et al., 2014) was used to analyze multivariate normality in the R environment for statistical computing (Version 3.2.4; R Core Team, 2016). Internal consistencies were calculated by R package psych (Version 1.6.9; Revelle, 2016).

Results

To adapt PID to the Argentinian population, the English PID items were forward-translated into and adapted to Argentinian Spanish by a receptive bilingual Argentinian psychologist (first author). Later, the items were back-translated into English by an independent bilingual Argentinian Professor of English. The back-translation was considered a close semantic approximation to the original English items. Consequently, the translated items were considered appropriate for inclusion in an Argentinian Spanish version of PID (see supplementary material). An example of item from the Preference for Intuition subscale is *My feelings play an important role in my decisions* and an example of item from the Preference for Deliberation subscale is *I prefer making detailed plans rather than leaving things to chance*.

The responses to the 18 PID items were analyzed for studying the score distribution. The Shapiro–Wilk statistics suggested that the score of each sample data item was drawn from a non-normally

distributed population (all the *p*s < .001). Moreover, Mardia's multivariate normality test results indicated that the data were not multivariate normal: Mardia's estimation of multivariate skewness = 45.28, skewness χ^2 = 1811.28, p < .001, Mardia's estimation of multivariate kurtosis = 405.13, kurtosis z = 13.03, p < .001.

An EFA based on a polychoric correlation matrix with extraction method of ordinary least squares and two oblique factor rotations was conducted. The parallel analysis based on minimum rank factor analysis on a polychoric matrix of correlations (Timmerman & Lorenzo-Seva, 2011) suggested retaining two factors, eigenvalues 4.36 and 3.82, and percentages of variance 27.4 and 24.4, respectively. Due to the variety of oblique rotations available, geomin and Crawford-Ferguson (CF) quartimax rotations were performed, as they would provide factor structures similar to those of the confirmatory factor analysis (CFA) results; and this technique was meant to be used later with the instrument's scores (Browne, 2001; Sass & Schmitt, 2010). The results of geomin and the Crawford–Ferguson (CF) equamax oblique rotation criteria presented strong coincidence with the expected intuitive and deliberative PID dimensions. Items that loaded on each factor were the same in both rotation criteria.

Nevertheless, the results of geomin and CF-equamax oblique rotation criteria showed that item 5, of the intuitive preference factors, resulted in a very low item-to-factor loading (0.02 on both rotation criteria); and item 6 resulted in a low item-to-factor loading and a large cross-loading (0.26 and 0.29, 0.25 and 0.28, respectively). In addition, the result of reliability analyses showed that alpha for PID-I was .77, but by eliminating item 5, the alpha value increased to .82; and alpha value for PID-D was .83. However, this value increased to .85 if item 6 was eliminated. Furthermore, those items were the only ones which, if absent, would result in an increment of the alpha values. Consequently, these items were dropped for the Argentinian Spanish PID.

Eight items corresponding to preference for intuition were retained because their item-to-factor loadings ranged from .52 to .83 for both rotation criteria, and the loadings on the deliberative factor were lower than |.18| for geomin, and lower than |.15| for CF-equamax. Eight items corresponding to preference for deliberation were retained because they presented item-to-factor loadings ranging from .54 to .79 for geomin rotation criterion, and .54 to .78 for CF-equamax, and the loadings on the intuitive factor were lower than |.15|, and |.14|, respectively. Although item 2 had presented problems in the validation of the English version, this item had .70 loading on the intuitive factor and almost zero cross-loading on the other factor. Internal consistencies for the PID Argentinian adaptation were Cronbach alpha = .85 for PID-D, and Cronbach alpha = .82 for PID-I.

Discussion

As hypothesized, the result of the PID psychometric analysis presented some weaknesses. Serious problems were found in items 5 and 6. Given that the scores of these items had already shown problems in other studies, we assumed that the difficulty of the items lay in their inability to relate to the corresponding factor. Accordingly, we decided to exclude these items from the Argentinian Spanish PID version.

STUDY 2

This study aimed to perform a CFA of the Argentinian Spanish PID version. Several studies on the CFA of PID are available: the original PID development, the Italian version, the English version, and a Slovak population (Betsch, 2004; Iannello, 2008; Witteman et al., 2009; Mikušková et al., 2015; respectively). These studies showed that PID presents a Comparative Fit Index (CFI) between .84 and .86, and a Root Mean Square Error of Approximation (RMSEA) between .06 and .09. Given that the cut-off values close to and above .95 for CFI and less than .07 for RMSEA are indicators of adequate fit, we assume that PID presented a less-than-adequate fit to data (Hooper et al., 2008; Hu & Bentler, 1999; Steiger, 2007).

Because PID does not have a simple structure and the results of the CFA of PID showed a fit to the data that can be considered less than good, we assumed that the CFA results of the 16-item PID model could present slightly better fit to data than the 18-item model, though without reaching a good fit.

Method

Participants. A convenience sample of 318 military academy undergraduate students (28 women) with a mean age of 22.5 years (*SD* = 2.6, age range of 18 to 29) was used for this study. The participants did not receive any compensation for their participation.

Materials and Procedure

Preference for Intuition and Deliberation scale (PID). The Argentinian Spanish version of PID (AE-PID) consists of 8 items that measure preference for intuition and 8 items that measure preference for deliberation. The items consist of statements, and the participants indicate their agreement using a 5-point scale ranging from 1 (*I very much disagree*) to 5 (*I very much agree*).

Procedures. An undergraduate sample was used to confirm the two-factor model of the adapted PID. To assess this model, a polychoric correlation was used with diagonally weighted least squares (DWLS) to estimate the CFA of the PID responses to the items, which is an ordered categorical scale (Yang-Wallentin et al., 2010). A series of fit indices were considered (Schweizer, 2010). For the goodness of fit of the two-factor model, three indices were used: CFI, Tucker-Lewis Index (TLI), and RMSEA. Cut-off values close to and above .95 were considered for CFI and TLI, and less than .07 for RMSEA (with the confidence interval upper limit of less than .08; Hooper et al., 2008; Hu & Bentler, 1999; Steiger, 2007). The R software package lavaan (Version 0.5-22; Rosseel, 2012) was used for the CFA.

Results

The Shapiro–Wilk statistics suggested that the score of each sample data item was drawn from a non-normally distributed population (all the *p*s < .001). Moreover, the results of the Mardia's multivariate normality test showed that data were not multivariate normal: Mardia's estimation of multivariate skewness = 32.29, skewness χ^2 = 1711.41, p < .001; Mardia's estimation of multivariate kurtosis = 338.95, kurtosis z = 18.92, p < .001.

The CFA results for the AE-PID model with 8 items for PID-I and 8 items for PID-D were Robust DWLSχ^2 (103) = 256.40, p < .05, normed χ^2 = 2.49, CFI = 0.902, TLI = 0.885, RMSEA = 0.069 (90% *CI* = 0.058 - 0.079).

The internal consistency for PID-I was .73, and for PID-D .75. If any items were removed, the internal consistency would not increase beyond the alpha value reached.

The modification indices were revised to find evidence of model misfit. Not fully unexpectedly, the three highest indices were associated with item 2 on preference for intuition. The path from latent factor preference for deliberation to item 2 exhibited the largest modification index value of 46.686 with an expected parameter change value of 0.373. In the original PID model, item 2 was specified as loading on preference for intuition factor, but the modification indices indicated that this item should additionally load on preference for deliberation factor. Although PID items should only target one factor, item 2 (*I listen carefully to my deepest feelings*), which was hypothetically related to preference for intuition, also seemed to tap into deliberative processing related to executive procedures for scrutinizing and examining one's mental contents. Although this item was included by Betsch (2004) in the preference for intuition, there seems to be statistical and content justification for a double-loading on the factors. The previous explanation together with the evidence of inverted loading of item 2 from the English validation of PID was the underpinning for modifying the model with this path freely estimated.

The CFA results for a respecified AE-PID model with item 2 double-loading on both factors were robust diagonally weighted least squares estimator (Robust DWLS) χ^2 (102) = 217.94, p < .05, normed χ^2 = 2.14, CFI = 0.926, TLI = 0.912, RMSEA = 0.060 (90% *CI* = 0.049 - 0.071). Consequently, the results related to the respecified PID model showed improvement in fit from the previously evaluated PID model with respect to all the indices.

Discussion

The AE-PID subscales showed acceptable internal consistency, and the original AE-PID scale yielded fit indices somewhat higher than those presented in previous research. Further, the respecified AE-PID model showed the best fit to the data. However, the fit of the respecified model was not acceptable per cut-off values. Therefore, analyses of measurement invariance across men and women and of latent means comparison could not be performed.

STUDY 3

Section 1: Developing a PID Model with good fit

The objective of this study was to develop a PID model with a good fit to data for more than one sample. This PID model would be appropriate for the study of measurement invariance and comparison of latent means across men and women.

Method

***Participants*.** The general population and student samples from Studies 1 and 2 were used. Merging the samples from Studies 1 and 2 resulted in a unified sample of 558 participants (148 women) with a mean age of 29.7 years (*SD* = 12.8, age range from 17 [one participant] to 89).

Materials and Procedure. AE-PID consisting of 8 items to measure preference for intuition and 8 items to measure preference for deliberation. The criteria for developing a putatively replicable model of PID was that the instrument should include at least four items per factor, the subscales internal consistencies should be equal to or greater than .70, and the items should load on a single factor. The factor indicators retained should structure a factorial model with a good fit consistent with the previously indicated CFA cut-off values. Accordingly, we repeatedly explored the goodness of fit of several models with both factors, but defined with different indicators, using in parallel the data from the samples of Studies 1 and 2. When we found evidence of lack of fit of models to data, we modified identical fragments of models and we studied their fit for samples 1 and 2, respectively. The R package lavaan was used for structural equation model analyses (Version 0.5-22; Rosseel, 2012).

Results. Different models with different indicators were explored iteratively for the AE-PID factors. The result of the selected AE-PID model was Robust DWLS χ^2 (47) = 77.66, $p < .01$, normed χ^2 = 1.65,CFI = .974, TLI = .963, RMSEA = 0.052, *ns* (90% CI = .030 - .072) for general population sample; and Robust DWLS χ^2 (47) = 76.42, $p < .01$, normed χ^2 = 1.63, CFI = 0.970, TLI = 0.958, RMSEA = 0.044, *ns* (90% CI = 0.025 - 0.062) for student sample. These results were interpreted as a good fit to the data. Internal consistencies for the general sample were PID-I alpha = .79, PID-D alpha = .79; for the student sample, PID-I alpha = .70, PID-D alpha = .70. The internal consistency of subscales did not increase by item elimination. The selected model was named revised AE-PID. It comprised items 4 (men *M* = 2.92, *SD* = 1.11; women *M* = 3.47, *SD* = 1.08), 8 (men *M* = 3.78, *SD* = 1.03; women *M* = 4.01, *SD* = 0.98), 9 (men *M* = 3.28, *SD* = 1.10; women *M* = 3.85, *SD* = 1.08), 12 (men *M* = 3.39, *SD* = 1.17; women *M* = 3.73, *SD* = 1.05), 15 (men *M* = 3.38, *SD* = 1.13; women *M* = 3.65, *SD* = 1.09),17 (men *M* = 3.19, *SD* = 1.06; women *M* = 3.62, *SD* = 1.15), and 18 (men *M* = 3.15, *SD* = 1.17; women *M* = 3.45, *SD* = 1.25) as indicator variables of preference for intuition, and items 1 (men *M* = 4.21, *SD* = 0.87; women *M* = 3.97, *SD* = 1.00), 7 (men *M* = 3.90, *SD* = 1.03; women *M* = 3.57, *SD* = 1.25), 11 (men *M* = 3.89, *SD* = 0.96; women *M* = 3.84, *SD* = 1.02), 14 (men *M* = 4.04, *SD* = 0.97; women *M* = 3.77, *SD* = 1.08) and 16 (men *M* = 3.31, *SD* = 0.94; women *M* = 3.26, *SD* = 1.10), as indicators of preference for deliberation (Argentinian Spanish items in Appendix; original English items; Betsch & Kunz, 2008). In order to obtain a better fit, we added a covariance between the error terms of the items related to preference for intuition 4 and 17, 4 and 18, 12 and 17, 15 and 18, 17 and 18 as well as those related to preference for deliberation 7 and 11. We included error covariance parameters in the model due to the modification indices and the item content overlap.

Discussion. The revised AE-PID model showed a good fit. The subscales of this revised model were shorter, but they exceeded the limit of 4 items by subscale and achieved an internal consistency equal to or greater than .70 for each of the subscales in the two samples. Consequently, the model was found to have satisfactory properties to evaluate the preference factors for intuition and deliberation. Eventually, we would be able to carry out the study of measurement invariance and latent means comparison in these factors.

Section 2: Latent means comparison

As mentioned, no studies are available on PID measurement invariance or on the difference in means of latent variables of preference for intuition or deliberation. We assume that, in part, this situation can be due to the lack of good fit to the data of the PID model. Therefore, we used the revised AE-PID model, which presents a good fit to the data, to study the invariance measurement and determine the difference in latent means between men and women.

On the one hand, by determining that a measuring instrument can achieve a strict measurement invariance, one may, with good reason, use the observed scores to study the difference in means of the subscales between groups. On the other hand, by determining the presence of differences between latent means, it may be deduced that comparisons between groups will not be affected by measurement errors derived from the observed scores.

Method

***Participants*.** The same sample as in Study 3 Section 1 was used.

Materials and Procedure. The revised AE-PID, consisting of 7 items to measure preference for intuition and 5 items to measure preference for deliberation. The measurement invariance test followed the guidelines proposed by Millsap and Yun-Tein (2004). The R package semTools (Version 0.4-11; semTools Contributors, 2016) was used for the invariance study and the R package lavaan was used for the latent means comparison (Version 0.5-22; Rosseel, 2012).

Results. Table 1 shows the result of the measurement invariance study across men and women. A strict measurement invariance across groups was reached. Thus, we used the observed scores to determine the difference in means between men and women. Women presented higher preference for intuition, $t(556) = -5.73$, $p < .001$, and lower preference for deliberation, $t(556) = 2.78$, $p < .01$, than men.

Model 5, which included mean invariance (along with configural, loading, threshold, and residual invariances), showed differences from Model 4, which did not include mean invariance. This indicates that there were differences in latent means in decision-making preferences across men and women.

In order to analyze the difference in latent means across men and women, we compared Model 3 with Model 1 (See Table 1; Thompson & Green, 2013). The scaled χ^2 difference test (Satorra, 2000) showed that the two models did not differ from each other, scaled $\chi^2(10.20) = 10.93$, $p = .38$. Given that a strong measurement invariance across men and women had been achieved, it was now possible to compare the latent means of preferences for intuition and for deliberation. The Wald z test results showed that the men presented less preference for intuition, $z = -4.627$, $p < .001$, but more preference for deliberation, $z = 2.058$, $p < .04$, than the women.

Table 1. Measurement invariance models across men and women.

Model	$\chi 2$	*df*	$\Delta\chi 2$	Δdf	*p*	CFI	ΔCFI	RMSEA	ΔRMSEA
1	116.38	94				.977		.044	
2	135.23	104	14.98	10	.13	.973	.004	.045	.001
3	174.24	138	39.12	34	.25	.967	.006	.043	.002
4	186.39	150	1.38	12	.58	.966	.001	.042	.001
5	411.86	152	2.70	2	< .001	.884	.082	.077	.035

Note: Model 1= Configural invariance, Model 2 = Factor loading invariance (weak invariance), Model 3 = Item threshold invariance (strong invariance), Model 4 = Measurement error variances invariance (strict invariance), Model 5 = Factor means invariance. Competing models are nested. *df* = degrees of freedom, CFI = Comparative Fit Index, RMSEA = Root Mean Square Error of Approximation.

For preference for intuition, the standardized effect size, latent *d,* was -.56, and for preference for deliberation, latent $d = .24$ (Thompson & Green, 2013). Following Funder and Ozer's interpretive guidelines (2019), after transforming *r*s to *d*s (Lenhard & Lenhard, 2016), adapted for latent means according to Hancock's proposal (2001), the standardized effect size can be considered more than very large for preference for intuition, and more than medium for preference for deliberation.

Discussion. The results show that the means of preference factors for deliberation and intuition are not equal across men and women in the population, and that the strength of the effects of women versus men on preference for intuition and for deliberation factors are more than very large and more than medium, respectively.

It is not surprising to find differences in preference for intuition in favor of women. However, the finding of a more than very large size of the standardized effect for preference for intuition has allowed us to calculate these differences with exactness, beyond the influence of measurement errors. In addition, the finding of a more than medium standardized effect size for preference for deliberation, i.e., less than large effect size, could explain why the literature has not reported the difference in preference for deliberation between men and women. Factors such as not large differences between men and women, a model with poor fit to the data, sample sizes, and/or measurement errors could have hindered the finding of differences between men and women in studies that based their comparisons on the variables observed.

GENERAL DISCUSSION

A sequence of studies have illustrated the conditions to determine in an accurate manner differences between men and women in the preference for intuition and deliberation. Our aim is that the procedure could serve as a model to establish differences between men and women in other variables.

Specifically, we have focused on the path to establish a PID model with a good fit, appropriate for studies of measurement invariance, included the adaptation to Argentinian Spanish of PID. Unlike other studies in the literature, this work provided details of the process followed to adapt PID to another language. Important issues were found for the PID. The psychometric analysis of the items showed, in agreement with the literature, that two items were weakly linked to their factors. Consequently, it was decided to eliminate these two items from the adapted version. This is the first time this criterion has been used for this instrument. The literature shows that these items were maintained although they were considered problematic. However, continuing to include PID items that are weakly linked to the factors would not help to improve the fit of the model to the data. The respecified AE-PID model, with item 2 loading on both factors, obtained a better fit. It was decided not to respecify this model further, in an attempt to maintain a balance between the received model and the respecified one in view of parsimony. Clearly, these steps highlight the psychometric issues of the PID. Moreover, a model with a good fit to the data was not available to make it possible to compare latent means between men and women, after successive comparisons of nested models in a study of factorial invariance. Finally, as a result of an iterative process of model selection, a revised AE-PID model with a good fit to data was selected from two different samples. This model allowed us to make the comparisons of latent means. Latent means for women in preference for intuition was higher than that for men, using the revised AE-PID. Notably, latent means for women was lower in preference for deliberation than that for men.

Although this study found differences in the preference for intuition and deliberation between men and women, its aim was not to demonstrate why such differences exist. Research on differences between men and women in decision making have attributed them to biological factors (such as hormones, e.g., Derntl et al., 2014; or brain areas, e.g., Sutterer et al., 2015) and/or social influences (e.g., Dorrough & Glöckner, 2019). However, future studies could determine the causal pathways to the differences in the preference for intuition or deliberation between men and women.

Limitations related to our adaptation studies

Some results of our research could be interpreted as having some possible limitations. We assume that the most serious limitation involved reusing the data obtained for Studies 1 and 2, to perform Study 3. This strategy poses the risk that it may capitalize on chance, and therefore, could put us in a better position to confirm our models. Consequently, upcoming AE-PID studies should include corroborating the findings of Study 3 with other samples. Another limitation of this research was not having access to datasets of previous studies using the PID. Such access may have allowed us to reach better results. In addition, we did not administer the PID together with instruments that evaluate similar constructs, such as the REI, as a point of reference to infer the degree of validity of our findings. A particular limitation of Study 2 was the low percentage of women participants. Another limitation was not forming men and women groups without age difference, which would have allowed us to rule out a putative age effect.

Conclusions

Determining the difference between men and women in psychological variables could need a series of steps on a measurement instrument development. As a result, latent means comparison detects differences between men and women without measurement errors. We expect showing this series of steps based on PID adaptation may serve as a general guide for researchers in order to make a methodologically sound comparison between women and men on a variable evaluated with a measurement instrument constituted of several items.

REFERENCES

Abós, Á., Burgueño, R., García-González, L., & Sevil-Serrano, J. (2021). Influence of internal and external controlling teaching behaviors on students' motivational outcomes in physical education: Is there a gender difference? *Journal of Teaching in Physical Education*. Advance online publication. https://doi.org/10.1123/jtpe.2020-0316

Archer, J. (2019). The reality and evolutionary significance of human psychological sex differences. *Biological Reviews*, *94*(4), 1381–1415. https://doi.org/10.1111/brv.12507

Betsch, C. (2004). Präferenz für Intuition und Deliberation (PID). Inventar zur erfassung von affekt- und kognitionsbasiertem entscheiden [Preference for Intuition and Deliberation (PID): An inventory

for assessing affect- and cognition-based decision-making]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *25*(4), 179–197. https://doi.org/10.1024/0170-1789.25.4.179

Betsch, C. (2008). Chronic preferences for intuition and deliberation in decision making: Lessons learned about intuition from an individual differences approach. In H. Plessner, C. Betsch, & T. Betsch (Eds.), *Intuition in judgment and decision making* (pp. 231–248). Lawrence Erlbaum Associates Publishers.

Betsch, C., & Iannello, P. (2010). Measuring individual differences in intuitive and deliberate decision making styles: A comparison of different measures. In A. Glöckner & C. Witteman (Eds.), *Tracing intuition: Recent methods in measuring intuitive and deliberate processes in decision making* (pp. 251–267). Psychology Press.

Betsch, C., & Kunz, J. J. (2008). Individual strategy preferences and decisional fit. *Journal of Behavioral Decision Making*, *21*(5), 532–555. https://doi.org/10.1002/bdm.600

Browne, M. W. (2001). An overview of analytic roation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*(1), 111–150. https://doi.org/10.1207/S15327906MBR3601_05

Caycho-Rodríguez, T., Vilca, L. W., Carbajal-León, C., White, M., Vivanco-Vidal, A., Saroli-Araníbar, D., Peña-Calero, B. N., & Moreta-Herrera, R. (2022). Coronavirus Anxiety Scale: New psychometric evidence for the Spanish version based on CFA and IRT models in a Peruvian sample. *Death Studies*, *46*(5), 1090–1099. https://doi.org/10.1080/07481187.2020.1865480

Colombo, B., Rodella, C., Riva, S., & Antonietti, A. (2013). The effects of lies on economic decision making. An eye-tracking study. *Research in Psychology and Behavioral Sciences*, *1*(3), 38–47.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). L. Erlbaum Associates.

Cools, E., & Van den Broeck, H. (2007). Development and validation of the cognitive style indicator. *The Journal of Psychology*, *141*(4), 359–387. https://doi.org/10.3200/JRLP.141.5.539-560

Derntl, B., Pintzinger, N., Kryspin-Exner, I., & Schöpf, V. (2014). The impact of sex hormone concentrations on decision-making in females and males. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00352

Dijkstra, K. A., van der Pligt, J., & van Kleef, G. A. (2017). Fit between decision mode and processing style predicts subjective value of chosen alternatives. *European Journal of Social Psychology*, *47*(1), 72–81. https://doi.org/10.1002/ejsp.2217

Dorrough, A. R., & Glöckner, A. (2019). A cross-national analysis of sex differences in prisoner's dilemma games. *British Journal of Social Psychology*, *58*(1), 225–240. https://doi.org/10.1111/bjso.12287

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Grevenstein, D., & Bluemke, M. (2022). Measurement invariance of the SOC-13 Sense of Coherence Scale across gender and age groups. *European Journal of Psychological Assessment*, *38*(1), 61–71. https://doi.org/10.1027/1015-5759/a000641

Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and mimic approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika*, *66*(3), 373–388. https://doi.org/10.1007/BF02294440

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*(1), 53–60. https://academic-publishing.org/index.php/ejbrm/article/view/1224

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. https://doi.org/10.1080/10705519909540118

Iannello, P. (2008). *Intuitive and analytical thinking in decision making: The role of mindreading and cognitive style in a strategic interactive context* [Unpublished doctoral thesis]. Università Cattolica del Sacro Cuore, Milan, Italy.

Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal*, *6*(2), 151–162. https://journal.r-project.org/archive/2014-2/korkmaz-goksuluk-zararsiz.pdf

Laborde, S., Dosseville, F., & Scelles, N. (2010). Trait emotional intelligence and preference for intuition and deliberation: Respective influence on academic performance. *Personality and Individual Differences*, *49*(7), 784–788. https://doi.org/10.1016/j.paid.2010.06.031

Lenhard, W., & Lenhard, A. (2016). Calculation of effect sizes. *Psychometrica* https://doi.org/10.13140/RG.2.2.17823.92329

Lorenzo-Seva, U., & Ferrando, P. J. (2013). FACTOR 9.2: A comprehensive program for fitting exploratory and semiconfirmatory factor analysis and IRT models. *Applied Psychological Measurement*, *37*(6), 497–498. https://doi.org/10.1177/0146621613487794

Mikušková, E. B., Hanák, R., & Čavojová, V. (2015). Appropriateness of two inventories measuring intuition (the PID and the REI) for Slovak population. *Studia Psychologica*, *57*(1), 63–82. https://doi.org/10.21909/sp.2015.01.674

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, *39*(3), 479–515. https://doi.org/10.1207/S15327906MBR3903_4

Miroshnik, K. G., Shcherbakova, O. V., & Kaufman, J. C. (2022). Kaufman Domains of Creativity Scale: Relationship to occupation and measurement invariance across gender. *Creativity Research Journal*, *34*(2), 159–177. https://doi.org/10.1080/10400419.2021.1953823

Monacis, L., Palo, V., de, Nuovo, S. D., & Sinatra, M. (2016). Validation of the rational and experiential multimodal inventory in the Italian context. *Psychological Reports*, *119*(1), 242–262. https://doi.org/10.1177/0033294116657623

Pachi, A., Sikaras, C., Ilias, I., Panagiotou, A., Zyga, S., Tsironi, M., Baras, S., Tsitrouli, L. A., & Tselebis, A. (2022). Burnout, depression and sense of coherence in nurses during the pandemic crisis. *Healthcare*, *10*(1), 134. https://doi.org/10.3390/healthcare10010134

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, *76*(6), 972–987. https://doi.org/10.1037/0022-3514.76.6.972

Phillips, W. J., Fletcher, J. M., Marks, A. D. G., & Hine, D. W. (2016). Thinking styles and decision making: A meta-analysis. *Psychological Bulletin*, *142*(3), 260–290. https://doi.org/10.1037/bul0000027

R Core Team. (2016). *R: A language and environment for statistical computing* (3.2.4) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Revelle, W. (2016). *psych: Procedures for personality and psychological research* (1.6.9) [Computer software]. Northwestern University. https://CRAN.R-project.org/package=psych

Richetin, J., Perugini, M., Adjali, I., & Hurling, R. (2007). The moderator role of intuitive versus deliberative decision making for the predictive validity of implicit and explicit measures. *European Journal of Personality*, *21*(4), 529–546. https://doi.org/10.1002/per.625

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research*, *45*(1), 73–103. https://doi.org/10.1080/00273170903504810

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in multivariate statistical analysis: A festschrift for Heinz Neudecker* (pp. 233–247). Springer Science + Business Media.

Schmitt, D. P. (2015). The evolution of culturally-variable sex differences: Men and women are not always different, but when they are…it appears not to result from patriarchy or sex role socialization. In T. K. Shackelford & R. D. Hansen (Eds.), *The Evolution of Sexuality* (pp. 221–256). Springer International Publishing. https://doi.org/10.1007/978-3-319-09384-0_11

Schmitt, T. A., & Sass, D. A. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement*, *71*(1), 95–113. https://doi.org/10.1177/0013164410387348

Schunk, D., & Betsch, C. (2006). Explaining heterogeneity in utility functions by individual differences in decision modes. *Journal of Economic Psychology*, *27*(3), 386–401. https://doi.org/10.1016/j.joep.2005.08.003

Schweizer, K. (2010). Some guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment*, *26*(1), 1–2. https://doi.org/10.1027/1015-5759/a000001

Scott, S. G., & Bruce, R. A. (1995). Decision-taking style: The development and assessment of a new measure. *Educational and Psychological Measurement*, *55*(5), 818–831. https://doi.org/10.1177/0013164495055005017

semTools Contributors. (2016). *SemTools: Useful tools for structural equation modeling. R package* (0.4-11) [Computer software]. http://cran.r-project.org/package=semTools

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, *42*(5), 893–898. https://doi.org/10.1016/j.paid.2006.09.017

Stevenson, S. S., & Hicks, R. E. (2016). Trust your instincts: The relationship between intuitive decision making and happiness. *European Scientific Journal*, *12*(11), 463–483. https://doi.org/10.19044/esj.2016.v12n11p463

Sutterer, M. J., Koscik, T. R., & Tranel, D. (2015). Sex-related functional asymmetry of the ventromedial prefrontal cortex in regard to decision-making under risk and ambiguity. *Neuropsychologia*, *75*, 265–273. https://doi.org/10.1016/j.neuropsychologia.2015.06.015

Tateneni, K., Mels, G., Cudeck, R., & Browne, M. W. (2009). *CEFA: Comprehensive Exploratory Factor Analysis* (3.04) [Computer software]. https://osu.app.box.com/s/ulmj5kh0mzpof7hy3ith7ksjbvv8gmdg

Thompson, M. S., & Green, S. B. (2013). Evaluating between-group differences in latent variable means. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 163–218). Information Age Publishing.

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, *16*(2), 209–220. https://doi.org/10.1037/a0023353

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. https://doi.org/10.1177/109442810031002

Wang, Y., Highhouse, S., Lake, C. J., Petersen, N. L., & Rada, T. B. (2015). Meta-analytic investigations of the relation between intuition and analysis. *Journal of Behavioral Decision Making*, *30*(1), 15–25. https://doi.org/10.1002/bdm.1903

Witteman, C., van den Bercken, J., Claes, L., & Godoy, A. (2009). Assessing rational and intuitive thinking styles. *European Journal of Psychological Assessment*, *25*(1), 39–47. https://doi.org/10.1027/1015-5759.25.1.39

Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling*, *17*(3), 392–423. https://doi.org/10.1080/10705511.2010.489003

Youssef-Morgan, C. M., van Zyl, L. E., & Ahrens, B. L. (2022). The Work Gratitude Scale: Development and evaluation of a multidimensional measure. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/article/10.3389/fpsyg.2021.795328

CRediT AUTHORSHIP CONTRIBUTION STATEMENT

Alejandro César Cosentino: Conceptualization; Data Curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing - Original Draft; Writing - Review & Editing. Susana **Celeste Azzollni**: Funding acquisition; Investigation; Project administration; Resources; Supervision.

ACKNOWLEDGMENTS

We appreciate the collaboration of Dr. Alejandro Castro Solano. We appreciate the contribution of Lic. Sergio Edgardo Castillo in his supportive role as translator.

History of the manuscript

Received	18/06/2021
Accepted	08/06/2022
Published (online)	20/12/2022
Published	03/07/2023