Journal of Business Research 156 (2023) 113539

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

Multigroup analysis of more than two groups in PLS-SEM: A review, illustration, and recommendations

Jun-Hwa Cheah [a,b], Suzanne Amaro [c,*], José L. Roldán [d]

[a] Norwich Business School, University of East Anglia, UK
[b] School of Business and Economics, Universiti Putra Malaysia, Selangor, Malaysia
[c] Higher School of Technology and Management, Polytechnic Institute of Viseu, Campus Politécnico, 3504-510 Viseu, Portugal
[d] Department of Business Administration and Marketing, Universidad de Sevilla, Spain

ARTICLE INFO

Keywords:
Multigroup Analysis
Group Comparison
Non-parametric Distance-Based Test
Non-parametric Permutation-Based Test
Omnibus Test Group Differences
Partial Least Squares Structural Equation Modeling

ABSTRACT

Multigroup analysis (MGA) in partial least squares structural equation modeling (PLS-SEM) has grown considerably in the past few years in many different research fields, particularly in the business area. However, a close examination of MGA in PLS-SEM articles revealed much less research that compared more than two groups. Furthermore, research applying MGA in PLS-SEM with more than two groups has several constraints. For instance, most researchers need clarification about using either the omnibus test of group differences (OTG) or non-parametric distance-based tests (NDT) for an overall difference across the groups. Moreover, they do not handle family-wise error when comparing more than two groups, nor do they check for measurement invariance. This article uses an empirical illustration to fully understand multigroup analysis with more than two groups, providing valuable guidelines and comprehensible recommendations for researchers applying PLS-MGA.

1. Introduction

Group comparisons have been conducted in many research fields recently because they are of great interest. Indeed, researchers can uncover differences in the subgroups within the entire population that are not visible when examining the entire sample (Matthews, 2017). For example, it provides a better understanding of consumer behavior in the marketing field, which is paramount for marketers to develop marketing strategies and provide superior value. In many real situations, the assumption of homogeneity is unrealistic because individuals, groups, or organizations are likely to be heterogeneous regarding their perceptions and evaluations of latent constructs (Sarstedt & Ringle, 2010). This is especially true for business research, which often examines differences in parameters related to distinct subpopulations, such as cultures and countries (Sarstedt et al., 2011; Ting et al., 2019). Notably, when performing on an aggregated data level, ignoring population heterogeneity can seriously bias the results and yield inaccurate management conclusions (Becker et al., 2022; Hair et al., 2019; Schlägel & Sarstedt, 2016). These arguments clearly illustrate the value and need of conducting group comparisons.

Multigroup analysis (MGA) is an approach that has been broadly used for group comparisons. It is a set of advanced techniques that are usually applied when researchers want to examine differences between categorical variables (i.e., gender or countries) (Hair et al., 2018) or continuous variables that can be categorized through a dichotomization process or cluster analysis (Hair et al., 2019). MGA can be executed in partial least squares structural equation modeling (PLS-MGA), and researchers can test for meaningful differences between the structural paths of multiple groups (Matthews et al., 2018).

Despite several studies with guidelines for applying PLS-MGA (e.g., Cheah et al., 2020; Hair et al., 2018; Matthews, 2017; Sarstedt et al., 2011), the examples provided are MGA with only two groups. However, researchers frequently encounter situations where they would like to compare more than two groups (Hair et al., 2018; Sarstedt et al., 2011).[1] To date, empirical research with more than two groups appears to be rarer. Nevertheless, specific details must be addressed when comparing more than two groups that have not been adequately evidenced in existing research. In fact, many business researchers still fail or are unaware to fully utilize the suitability evaluation procedure when conducting MGA with more than two groups in PLS-SEM. For instance,

* Corresponding author.
E-mail addresses: jackycheahjh@gmail.com (J.-H. Cheah), samaro@estgv.ipv.pt (S. Amaro), jlroldan@us.es (J.L. Roldán).

[1] Researchers have expressed a strong desire on Internet forums such as https://www.smartpls.de, to clarify how multigroup analysis can be performed for more than two groups, particularly within a PLS path modelling framework.

https://doi.org/10.1016/j.jbusres.2022.113539
Received 11 April 2022; Received in revised form 2 December 2022; Accepted 3 December 2022
Available online 15 December 2022

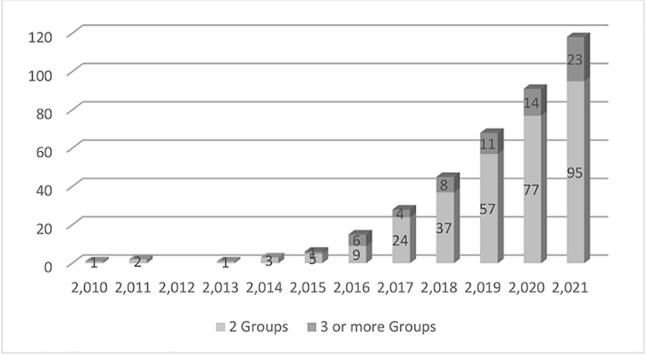


Fig. 1. Number of articles published using PLS-SEM MGA by years and groups.

when conducting PLS-MGA with more than two groups, it is necessary to conduct all possible pairwise group comparisons using the non-parametric permutation-based test (NPT). This increases the overall probability of a Type I error (also known as the familywise error rate), i.e., the likelihood of finding a significant result is higher when it is not (Hair et al., 2018). Hair et al. (2018) recommend performing the Bonferroni correction or the Šidák procedure to control this type of error when using NPT.

Another recommendation is to calculate the omnibus test of group differences (OTG) to examine the overall difference in parameters across multiple groups before conducting pairwise group comparisons (Hair et al., 2018). According to Hair et al. (2018), "If the OTG approach indicates a significant effect, we can conclude that the path coefficient of at least one group differs significantly from those of the other groups. However, this does not mean that all the groups' path coefficients differ significantly" (p.157). Therefore, the OTG emphasizes estimating a single parameter rather than the entire or complete model.

Furthermore, a more recent and robust technique could also be performed, namely, the non-parametric distance-based test (NDT) that compares the model-implied correlation matrix across two (or more) groups using a permutation procedure (Klesel et al., 2019; Klesel et al., 2022). In particular, the NDT technique suggested by Klesel et al. (2019) compares all parameters simultaneously between groups. Indeed, "all model parameters are taken into account, and thus, the complete model is compared across groups" (Klesel et al., 2022, p. 27). Thus, the aim is to explore whether the proposed research models or theories behave differently across groups. Therefore, the OTG might show a significant difference in one parameter (i.e., a specific path) while the NDT might not indicate any significant differences because it assesses the whole model and not only one path.

Recognizing the paucity of research on MGA with more than two groups, the first objective of this research is to examine articles that have used PLS-MGA to determine how much research has been conducted with more than two groups and whether the recommendations of Hair et al. (2018) and Klesel et al. (2019) are being followed. The second objective of this paper is to illustrate and provide appropriate guidelines for PLS-SEM with more than two groups. Specifically, it presents an empirical example comparing three groups (countries) with all the steps necessary to perform MGA in PLS-SEM, providing researchers with a valuable basis to conduct and present the results when performing MGA with more than two groups in PLS-SEM. This is crucial to increase the usage of MGA in PLS-SEM and to disseminate rigorous methodological standards in business domains such as information systems, human resource management, and marketing, among others.

Table 1
Reporting in PLS-MGA studies with three of more groups.

Test	Number of studies reporting
Measurement Invariance using MICOM	21 (31.34 %)
OTG	3 (4.48 %)
NDT	0 (0 %)
Šidák	1 (1.49 %)
Bonferroni	1 (1.49/%)

2. Research applying PLS-SEM multigroup analysis

Research in PLS-SEM is difficult to confine to specific disciplines and can be found in many different research fields. Therefore, in March 2022, two general databases, Science Direct and SCOPUS, were used to search for articles on PLS-SEM multigroup analysis.

The literature search was based on the keywords "PLS-SEM Multigroup Analysis", "PLS-SEM Multigroup", "PLS-MGA", and "PLS Multigroup" within the article title, abstract, and keywords. The full text of each article was reviewed to eliminate those that did not perform PLS-MGA. Only published empirical research articles were considered (abstracts, conceptual papers, book chapters, conference papers, and dissertations were excluded). The search was limited until 2021. A total of 378 articles conducted a PLS-SEM multigroup analysis in 183 different journals. The distribution by years and the number of compared groups are provided in Fig. 1.

It should be noted that articles that conducted PLS-MGA but did not mention this in the title, abstract, or keywords do not appear in the results. Indeed, while the results indicate that the first article published applying PLS-SEM MGA appeared in 2010, the first paper was published by Keil, Tan, Wei, Saarinen, Tuunainen, and Wassenaar in 2000. Although some papers may not have been considered, the literature search carried out is very significant for the purpose of this study. It reveals the exponential growth in the number of PLS-MGA articles in the last few years. Of the 378 articles using PLS-SEM MGA, only 67 (18 %) compare three or more groups. In fact, many more studies have used PLS-SEM to compare two groups in the literature.

Some of the most popular two-group comparisons have been gender, with 21 % of the articles (e.g., Ghasemy et al., 2020; Lim et al., 2021; Velayutham et al., 2012) and country (11 % of articles) (e.g., Ahmad et al., 2021; Ramli et al., 2019). However, many other examples can be found, from comparing mobile coupon users between novices and experts (e.g., Carranza et al., 2020) to comparing the effect of corporate social responsibility on economic performance between family and non-family businesses (e.g., Yáñez-Araque, et al., 2021). Regarding more than two-group comparisons, 25 % of the articles have compared countries (e.g., Basco et al., 2020; Roy et al., 2018), and 11 % compared age groups (e.g., Mahmoud et al., 2020; Schade et al., 2016).

Each article was carefully reviewed to see whether measurement invariance between groups had been assessed. In the case where three or more groups were compared, the article was further examined to check if the OGT or NDT approaches had been applied and the Bonferroni or Šidák procedures.

Among the 378 articles, 61 % (n = 229) did not mention whether measurement invariance between all groups had been assessed. This is an essential issue because measurement invariance is necessary to ensure the validity of the results, i.e., that group differences are not due to different meanings of variables in groups or the measurement scale (Hair et al., 2018; Henseler et al., 2016). It was only in 2011 that Sarstedt et al. recommended considering measurement invariance. One year later, Chin et al. (2012) introduced permutation-based multigroup invariance to assess measurement invariance in PLS-SEM. However, this technique was not fully developed then because it only allowed researchers to assess configural invariance and use differences in group parameters for measurement loadings to determine full or partial invariances. Henseler et al. (2016) advanced measurement invariance testing by developing the measurement invariance of composite models (MICOM) based on the permutation technique. This technique caters to both common-factor and composite models. It deals with configural invariance, compositional invariance, equal variances, and equal means, providing a clearer understanding of achieving partial and full

invariance (Hair et al., 2018). With this proven technique, research using the measurement invariance significantly increased. Yet, among the articles published after the MICOM technique was published, more than 61 % of the articles still did not report measurement invariance. This supports Sarstedt et al., (2022a) claim that few studies implement the MICOM technique when investigating the categorical moderators' impact using MGA.

Considering the 67 studies that performed MGA in PLS-SEM with three or more groups, only 31.34 % reported measurement invariance using MICOM (Table 1), and only three performed the OTG test. Furthermore, only two conducted the Bonferroni correction or the Šidák procedure to handle the family-wise error. Moreover, no articles conducted the NDT. Therefore, most PLS-MGA research with three or more groups does not follow Hair et al. (2018) or Klesel et al. (2019) recommendations.

Another interesting fact from published MGA research is that many PLS-SEM users compare only two groups when comparing more would make more sense. For instance, researchers divide the sample into two age groups, despite having the data to divide the sample into more age groups to perform comparisons (e.g., Assaker et al., 2015; Trojanowski & Kułak, 2017). On the other hand, some articles claiming to have conducted PLS-MGA were eliminated from the analysis because they only tested the model using different groups but did not test for significant differences between the groups. Therefore, this review further confirms the need to provide an empirical example of PLS-MGA in general and, in particular, with three or more groups.

3. Techniques to conduct multigroup analysis with more than two groups

Looking back at the PLS-SEM literature, there are two prominent techniques to perform multigroup analysis with more than two groups, namely the (i) *Omnibus Test of Group (OTG)* and the (ii) *Non-parametric Distance-Based Test (NDT)*. Sarstedt et al. (2011) introduced the OTG to compare a single parameter across more than two groups. This test incorporates both the bootstrapping and permutation techniques to generate a criterion similar to the overall F test. Although this approach is comparable to an analysis of variance (ANOVA) F-test for analyzing more than two groups, it retains the type I error level (familywise error rate) and offers an appropriate statistical power without relying on distribution assumptions (Bortz, Lienert, & Boehnke, 2003; Pitman, 1938). Furthermore, if the OTG technique shows a significant effect, this indicates that the path coefficient of at least one group differs significantly from those of the other groups.

There are, however, two limitations to this approach. First, OTG might not be sufficient to detect group differences because this technique consistently reveals significant parameter differences, mainly when the number of bootstrap runs increases in the estimation (Cheah et al., 2020). Second, the approach only looks at whether the average bootstrap estimates of a parameter differ significantly between groups but not the estimated parameters (Klesel et al., 2022). Therefore, when the number of bootstrap runs increases, the value of the F-test statistic that compares the bootstrap means across groups will also increase as the within-group variation, i.e., the variance of the bootstrap means reduces, even if there is no group difference in the population.

Conversely, the F-test statistics depend on the permutation samples used to estimate the reference distribution under the null hypothesis that no group differences are relatively trivial and, thus, closely distributed right of zero. Therefore, Klesel et al. (2022) have argued that OTG is not recommended because the procedure detects differences between groups that may not exist. Furthermore, the OTG results do not provide conclusive evidence whether there are specific differences in the path coefficients between the groups. Thus, it is necessary to apply a pairwise comparison test, which will be explained in the following section.

This limitation has encouraged Klesel et al. (2019) to develop the NDT approach as an alternative approach to compare the differences between two or more groups in a complete model rather than a single parameter. The NDT can help researchers compare all parameters simultaneously between groups. First, the average distance between the model-implied indicator correlation matrices is examined to assess whether the complete model differs across two or more groups. The test compares the model-implied indicator correlation matrix between groups using either the average squared Euclidean distance (dL) or the average geodesic distance (dG). Thus, the complete structural model can be compared across groups (see Klesel et al., 2022). In other words, the test investigates the distances between the indicators model-implied correlation matrices across groups based on the proposed model.

Furthermore, the NDT approach can compare the correlation matrix of indicators for common factors and composites (regardless of loadings and weights) and measurement error correlations. In particular, the NDT employs permutation to acquire the reference distribution of the distances (Klesel et al., 2019). Therefore, if the group difference does not exist in the population's model-implied indicator correlation matrix, both distances should be closely distributed around zero within the bounds of sampling variation. In other words, the two groups differ if the distance exceeds zero. On the contrary, if the model-implied correlation matrix differs between groups, the differences should be significantly greater than zero, and the test is most likely to indicate a significant result (Klesel et al., 2022).

If the NDT result is significantly established, further steps, such as pairwise comparisons, must be conducted to investigate the differences in more depth, e.g., specific path coefficients, regardless of direct or indirect effect comparisons (Cheah et al., 2021). In other words, specifying specific parameters for comparison must be implemented after the NDT result. Hence, researchers should first identify the path coefficients based on clear theoretical linkages before using a bucket list of MGA statistical approaches, such as parametric test equal variances (PTE; see Keil et al., 2000), parametric test unequal variances (PTU; see Sarstedt et al., 2011), non-parametric bootstrap-based test (NBT; see Henseler, 2012), the non-parametric permutation-based test (NPT; see Chin, 2003; Chin & Dibbern, 2010), confidence intervals overlap (CIO; see Sarstedt et al., 2011), and confidence intervals cover parameter (CIP; see Sarstedt et al., 2011).

Evidently, Klesel et al.'s (2019; 2022) NDT technique can also be used to compare a single parameter between two groups and a complete structural model. However, among these approaches, Klesel et al.'s (2022) simulation study recommends using NPT when testing a single parameter. The reason is that NPT demonstrated higher power while keeping the predetermined significance threshold compared to other techniques developed to compare a single parameter. With this supporting evidence, this study compares a specific parameter between groups using NPT as the most robust technique.

As the name implies, NPT uses a permutation test to obtain the reference distribution of the difference in the parameters from which the critical values are estimated. It is used to determine whether a difference in parameters is statistically significant. The pairwise comparison (similar to a *t*-test) only allows researchers to compare two groups simultaneously. Therefore, researchers must run this test multiple times if there are more than two groups. For example, ten comparisons must be performed to compare a specific path relationship across five groups.

To avoid alpha inflation in the MGA test of more than two groups, researchers must conduct all plausible pairwise group comparisons while controlling the familywise error rate (Type I error) using the p-value adjustment.[2] This can be easily addressed by using either the Bonferroni or Šidák corrections, which are often used and suggested in the PLS-SEM literature (Hair et al., 2018; Hair et al., 2019; Sarstedt &

[2] When conducting MGA with more than two groups, the p-value adjustment is used to avoid the familywise error rate. The purpose is to prevent a false conclusion in a series of hypothesis tests by means of avoiding the likelihood of making a Type 1 error.

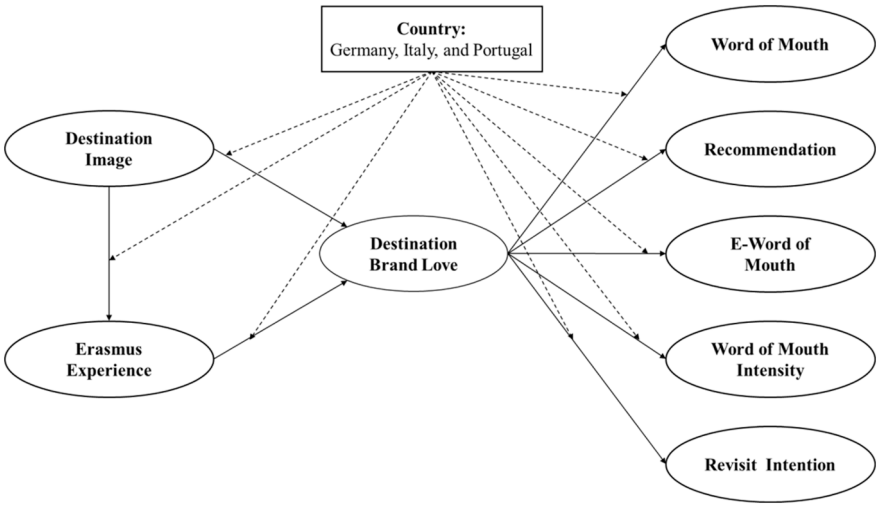


Fig. 2. Research Framework. Note: * means that it is crucial to adjust the conventional p-value with Šidák's or Bonferroni's adjustment when assessing the pairwise difference between groups.

Mooi, 2019).[3] The Bonferroni correction can be implemented using the formula alpha/m. For example, if researchers compare five groups, there would be m = 10 comparisons, producing a significance level of 0.05/10 = 0.005 instead of 0.05. Conversely, the Šidák correction employs a different formula, which is 1-(1-alpha)1/m. In other words, instead of using 0.05, a significant level of $1\text{-}(1\text{–}0.05)^{1/10} = 0.005116$ would be used for five groups with ten comparisons. Both p-value adjustments are conceivable, but researchers should be aware that the Bonferroni correction has lower statistical power than the Šidák correction in identifying pairwise group comparisons (direct and/or indirect effect) when the number of comparisons grows exponentially (see Hair et al., 2018).

4. Empirical illustration of a multigroup analysis with more than two groups: Procedures and guidelines

4.1. Model and data

The model used by Amaro et al. (2020) was replicated in this study to perform a multigroup analysis between more than two groups (Fig. 2). The sample used in their study to examine the antecedents and outcomes of Destination Brand Love (DBL) was composed of Erasmus students from three different countries. Germany, Italy, and Portugal. All measures for each item were rated on a five-point Likert scale.

4.2. Model estimation

Partial least squares structural equation modeling (PLS-SEM) was used to test for differences between Erasmus students regarding the antecedents and outcomes of Destination Brand Love in the three countries. PLS-SEM uses a causal-prediction approach that suits the prediction-oriented objective of the present research (Chin et al., 2020; Hair & Sarstedt, 2019; Sarstedt, Hair, & Ringle, 2022b). Importantly, PLS-SEM can estimate a very complex structural model (see Fig. 2), performing advanced assessments such as permutation tests, i.e., the OTG test and the NDT test (Hair et al., 2018).[4] To implement these advanced assessments, researchers should follow the five-step procedure shown in Fig. 3. The analysis was carried out to facilitate the explanation using SmartPLS version 3.3.9 (Ringle et al., 2015; Sarstedt & Cheah, 2019).

4.2.1. Step 1: Data preparation by generating data groups

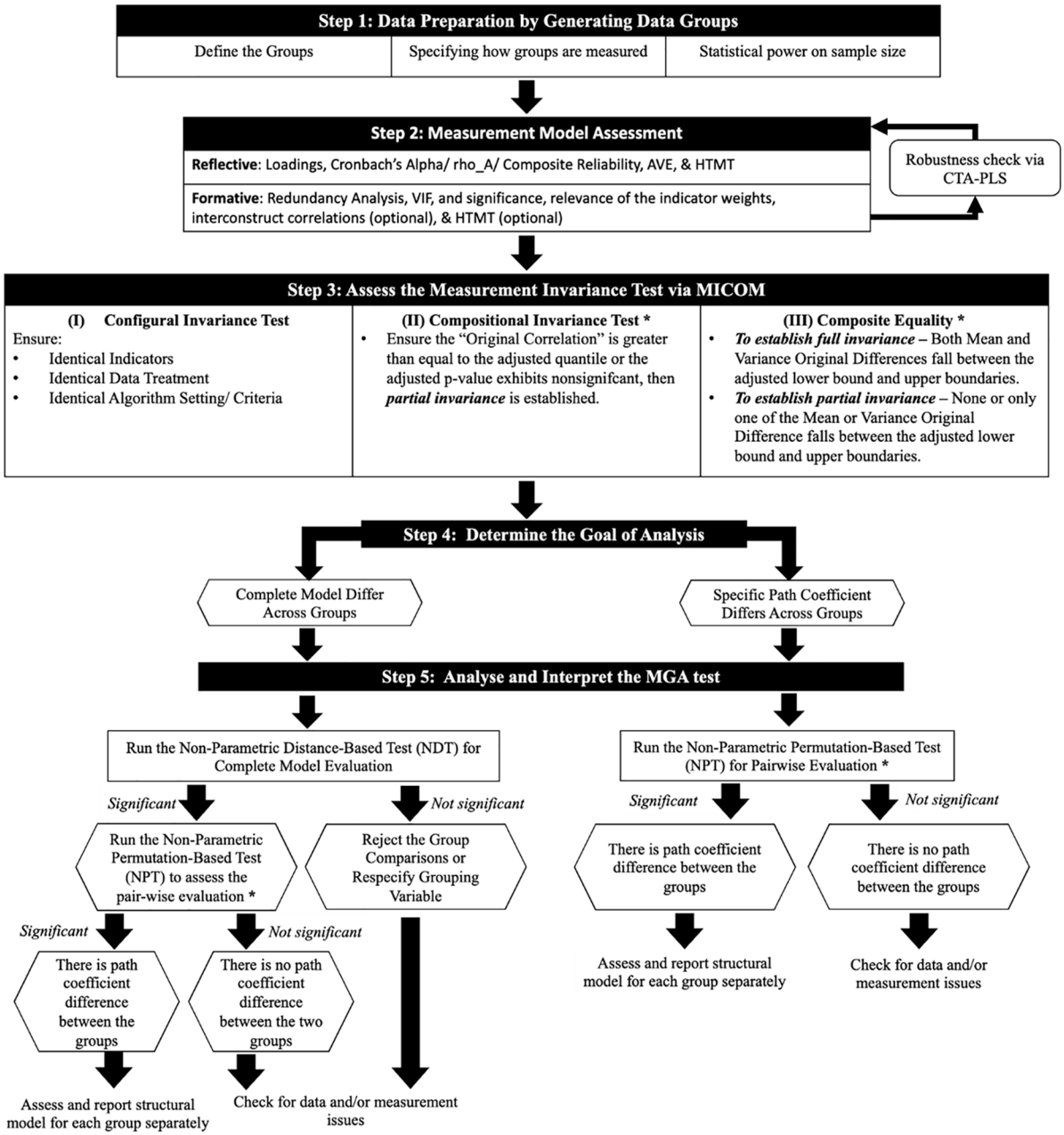
Data preparation is essential for ensuring rigor and avoiding misleading results in the group comparison study. There are three data preparation steps, namely (i) defining the groups, (ii) specifying how groups are measured, and (iii) statistical power on sample size.

First, several approaches can be used to define the groups for a comparison study. In particular, researchers frequently choose the categorical variable of interest from the dataset. In this case, prior knowledge of the theoretically plausible models is necessary to ensure a meaningful group comparison study. In other words, theory and observation are essential to generate the data groups. For example, suppose that both theory and observation show that Germany, Italy, and Portugal produce different results regarding the effects of Destination Image and Erasmus Experience on Destination Brand Love. In that case, researchers must position nationality as a moderator to examine overall relationships.

In addition, group comparisons could be conducted by drawing on a continuous moderator variable (e.g., switching costs or satisfaction). According to Hayes and Preacher (2014), artificial dichotomization through dividing points (i.e., percentiles) can be used when transforming continuous variables into categorical variables of more than two groups (low, medium, and high conditions). Hayes and Preacher (2014) stressed that estimations based on such transformed data should

[3] This study only focuses on using Bonferroni and Šidák corrections, which are commonly applied by empirical researchers in assessing the multiple pairwise comparisons of more than two groups (Hair et al., 2018; Sarstedt and Mooi, 2019). However, researchers must be aware that there are other correction methods that can be used for p-value adjustment. One of the well-established p-value adjustments that researchers can explore is Holm's p-value method. It is known to be less conservative than both Bonferroni and Šidák correction in dealing with familywise error (Holm, 1979). This correction employs stepwise adjustments to the significant level based on the rank order of the multiple tests' p-values (smallest to largest p-values). The formula to calculate Holm's correction is $\frac{\alpha}{(k-i+1)}$, in which αis often used as 0.05, k is the number of tests, and i is the rank number of pair tests based on the degree of significance (see Holm, 1979). This approach can also be executed on the following website: https://www.quantitativeskills.com/sisa/calculations/bonfer.htm.

[4] SmartPLS is used to estimate the measurement, structural model and MGA tests, such as measurement invariance of composite models and permutation test (Ringle, Becker, and Wende, 2015; Sarstedt and Cheah, 2019). In addition, this study also used the R statistical package (Hair et al., 2022b) like the cSEM package to perform the NDT approach (Rademaker & Schuberth, 2021) and the OTG approach in R (Sarstedt et al., 2011).



Note: * means that it is crucial to adjust the conventional p-value with Šidák's or Bonferroni's adjustment when assessing the pairwise difference between groups.

Fig. 3. Procedure for conducting MGA on two and more groups.

have strong underlying reasoning and justification. However, such an approach requires extra vigilance since it can reduce the statistical power to identify path differences and induce a downward bias in estimated path coefficients and group comparisons (Cheah et al., 2021). Hair et al. (2019) also stressed that artificial dichotomization through dividing points is an arbitrary and unscientific method of defining groups. The main reasons are that much of the information contained in the continuous moderator variable is lost during conversion, and it is unlikely that the split approach (either mean, median, or percentile) would result in an equal sample size. Therefore, this approach should not be used for rigorous research practices.

Another exciting and superior approach that several researchers have used to define groups is the FIMIX-PLS procedure (e.g., Mikalef et al., 2020; Sarstedt, Radomir, Moisescu, and Ringle, 2022c). It allows researchers to detect unobserved heterogeneity and define groups that are then used for multigroup analysis. Researchers have also used cluster analysis to define meaningful subgroups of objects (i.e., respondents, brands, products, or other entities) used for comparisons (see Arsenovic et al., 2021, Huaman-Ramirez & Merunka, 2019, Sanchez-Franco & Roldán, 2010). Cluster analysis is a multivariate data analysis technique that can maximize the homogeneity of groups within clusters while also maximizing the heterogeneity between groups (Hair et al., 2019; Hair, Page, and Brunsveld, 2020).

Second, it is necessary to specify how groups are measured. For

instance, researchers may be interested in comparing path coefficients across more than two existing groups, such as countries, as in the example presented in this study (i.e., Germany vs Italy, Germany vs Portugal, and Italy vs Portugal). Germany could be categorized as group 1, Italy as group 2, and Portugal as group 3 to distinguish and detect the output displayed in any PLS-SEM software or R statistical environment. In contrast, if a continuous variable (e.g., switching costs) is used to investigate group comparisons of the low, medium, and high conditions, then the percentile transformation can be an option. In this case, the low condition will represent group 1, the medium condition will represent group 2, and the high condition will represent group 3. For more advanced group comparisons of more than two groups, researchers can use combinations of categorical variables to compare more than two groups (e.g., nationality: Germany, Italy, and Portugal and generation: Gen-X, Gen-Y, and Gen-Z), which creates more complex multiple outcome groups.

Third, examining the sample size is crucial after specifying the groups that will be compared. When comparing more than two groups, it is necessary to ensure that each subgroup is large enough and comparable in size to detect the moderating effect (Hair et al. 2022a). This notion is consistent with Aguinis et al.'s (2017) view that it would be unreliable to estimate moderator effects with unequal sample sizes across subgroups because it would reduce statistical power and lead to underestimation of moderating effects, even if the total sample sizes are substantially large. To mitigate this problem, researchers should select similar sample sizes for each group to maximize sample variance (see Aguinis et al., 2017). However, although oversampling a smaller group of respondents can be advantageous to increase statistical power, it can also result in a misrepresentation of the smaller group in the sampling frame concerning the actual study population (Becker et al., 2013).

There are several approaches to determine the minimum sample size recommended in PLS-SEM. The standard practice recommends a statistical power of 80 % when assuming a significance level of 5 % for each subgroup, as Cohen (1988) and Hair et al. (2022a) recommend. Therefore, groups with fewer observations than those recommended for a statistical power of 80 % should be avoided with PLS-MGA unless it is impossible in a situation where the population being sampled is small. For example, business-to-business (B2B) research or team-based assessments in organizational research generally have a small population (Hair et al., 2019). Thus, these samples will be smaller when low response rates occur during data collection and after data cleaning.

Furthermore, researchers could look at the work of Aguirre-Urreta and Rönkkö (2015) to determine the power-based sample size through R programming. Alternatively, it is also possible to determine the minimum sample size needed for a path model using the gamma-exponential or inverse square root (Kock and Hadaya, 2018). For example, the gamma exponential method requires a minimum sample size of 146. In contrast, the inverse square root requires a minimum sample size of 160, especially if researchers are not aware of the path coefficient with the minimum absolute magnitude. Although both approaches occasionally result in minimal overestimation, this minor flaw guarantees that the problem of undetected moderation due to inadequate power is avoided.

4.2.2. Step 2: Measurement model assessment

Data were analyzed and interpreted using the two-step approach: assessment of the measurement model and the structural model (Hair et al., 2020; Hair et al. 2022a).[5] In assessing the measurement model, the relevant criteria for reflective and formative constructs differ (cf. Hair et al. 2022a for rules of thumb). For a reflective measurement mode, indicator loadings should be significant, with a value of at least 0.708. Additionally, measurements must be valid (i.e., convergent validity: average variance extracted, and discriminant validity: heterotrait-monotrait (HTMT) ratio correlation) and reliable (i.e., indicator reliability, Cronbach's alpha, rho A, and composite reliability). For formative measurement modes, Hair et al. (2022a) recommend assessing convergent validity and to check for multicollinearity issues, using the variance inflation factor. Finally, they also recommend examining each indicator's outer weight's significance. In order to ensure more robust construct validity of formative measurement modes, Urbach and Ahlemann (2010) suggested assessing discriminant validity using the interconstruct correlations between formative constructs and all other constructs. Interconstruct correlations lower than 0.7 indicate sufficient discriminant validity (Bruhn et al., 2008; MacKenzie et al., 2005). In addition, the latent variable scores of the formative construct can be used in the HTMT assessment to evaluate the discriminant validity with other constructs. Notably, although measurement models were often theoretically substantiated, researchers can depend on confirmatory tetrad analysis (CTA-PLS) to avoid potential model misspecification (see Gudergan et al. 2008). Finally, if the measurement models meet all the criteria, the evaluation of the measurement invariance test in Step 3 can be carried out.

The reflective constructs in this study were assessed first.[6] In particular, factor loadings, Cronbach's alpha (CA), rho A, composite reliability (CR), and average variance extracted (AVE) were used to evaluate the constructs' internal consistency and convergence validity (complete and split datasets from Germany, Italy, and Portugal) (Hair et al. 2022a). Table 2 shows that all item loadings exceed the recommended value of 0.5, except for two items, which were BL2_rev and EWM3. Therefore, these items were removed due to low loading for the three datasets (Hair et al. 2022a). As a result, the CA, rho_A, and CR values obtained afterward are larger than 0.7, and the AVE values are greater than the threshold value of 0.5 (Hair et al. 2022a; Nunnally & Bernstein, 1994). Thus, the complete and split samples of the three groups demonstrated a satisfactory result in terms of convergent validity and internal consistency.

Subsequently, the formative construct of destination image was assessed. First, all items of destination image were checked for collinearity issues.[7] Table 3 shows that the variance inflation factor (VIF) values are lower than 3, indicating that all formative items are distinct. Second, the outer weights and significance of each item were examined.

[5] Before assessing MGA of more than two groups, the measurement model assessment can be conducted by two different approaches of confirmatory composite analysis (CCA) from Hair et al. (2020) or Henseler and Schuberth (2020). CCA is a sequence of steps that can be performed with a composite-based SEM method (i.e., PLS-SEM) to confirm reflective and formative measurement models – particularly on ensuring reliability and validity – within a specific nomological network. The choice or usage of CCA is deeply intertwined between the work of Hair et al. (2020) and Henseler and Schuberth (2020). Therefore, one should read both studies to understand the purpose of using CCA and the distinction between different perspectives on CCA.

[6] Revisit Intention is treated as a single item. Since a single-item construct is equal to its measure, the indicator loading, conventional reliability, and convergent validity assessment will result in a value of 1.00, making the interpretation of the measurement difficult.

[7] In formative measurement model evaluation, redundancy analysis is not being assessed because neither a single global item nor reflective construct that measures the same concept is being collected in this study (see Cheah et al., 2018 on the implementation of redundancy analysis).

Table 2
Assessment of Reliability and Convergent Validity.

Dataset	Construct	Item	DBL	CA	rho_A	CR	AVE	FC
Complete (all countries combined) n = 2237	Brand Love	BL1	0.833	0.912	0.923	0.923	0.669	2.818
		BL2_rev	*					
		BL3	0.878					
		BL4	0.888					
		BL5_rev	0.608					
		BL6	0.842					
		BL7	0.826					
	Erasmus Experience	EEX1	0.835	0.844	0.865	0.890	0.622	1.798
		EEX2	0.864					
		EEX3	0.875					
		EEX4	0.692					
		EEX5	0.648					
	E-WoM	EWM1	0.879	0.758	0.772	0.892	0.804	1.180
		EWM2	0.914					
		EWM3	*					
	Recommend	RCM1	0.881	0.888	0.891	0.931	0.818	2.715
		RCM2	0.906					
		RCM3	0.924					
	Revisit Intention	RVST1	1.000	1.000	1.000	1.000	1.000	1.579
	WoM	WM1	0.894	0.877	0.879	0.924	0.803	2.698
		WM2	0.916					
		WM3	0.878					
	WoM Intensity	WMI1	0.725	0.813	0.823	0.877	0.641	1.885
		WMI2	0.845					
		WMI3	0.839					
		WMI4	0.787					
Germany n = 921	Brand Love	BL1	0.821	0.898	0.910	0.922	0.667	2.745
		BL2_rev	*					
		BL3	0.876					
		BL4	0.886					
		BL5_rev	0.623					
		BL6	0.842					
		BL7	0.824					
	Erasmus Experience	EEX1	0.828	0.840	0.859	0.888	0.616	1.792
		EEX2	0.870					
		EEX3	0.863					
		EEX4	0.697					
		EEX5	0.638					
	E-WoM	EWM1	0.877	0.787	0.837	0.902	0.821	1.174
		EWM2	0.935					
		EWM3	*					
	Recommend	RCM1	0.880	0.876	0.880	0.924	0.801	2.559
		RCM2	0.893					
		RCM3	0.912					
	Revisit Intention	RVST1	1.000	1.000	1.000	1.000	1.000	1.517
	WoM	WM1	0.892	0.875	0.876	0.923	0.800	2.589
		WM2	0.919					
		WM3	0.871					
	WoM Intensity	WMI1	0.728	0.811	0.819	0.876	0.639	1.867
		WMI2	0.831					
		WMI3	0.848					
		WMI4	0.785					
Italy n = 761	Brand Love	BL1	0.845	0.899	0.912	0.923	0.670	2.895
		BL2_rev	*					
		BL3	0.878					
		BL4	0.878					
		BL5_rev	0.615					
		BL6	0.839					
		BL7	0.825					
	Erasmus Experience	EEX1	0.827	0.842	0.860	0.889	0.618	1.832
		EEX2	0.856					
		EEX3	0.871					
		EEX4	0.698					
		EEX5	0.655					
	E-WoM	EWM1	0.860	0.737	0.768	0.882	0.790	1.210
		EWM2	0.916					
		EWM3	*					
	Recommend	RCM1	0.876	0.890	0.892	0.932	0.819	2.685
		RCM2	0.910					
		RCM3	0.929					
	Revisit Intention	RVST1	1.000	1.000	1.000	1.000	1.000	1.585
	WoM	WM1	0.893	0.872	0.875	0.921	0.795	2.758

(*continued on next page*)

Table 2 (*continued*)

Dataset	Construct	Item	DBL	CA	rho_A	CR	AVE	FC
		WM2	0.903					
		WM3	0.879					
	WoM Intensity	WMI1	0.724	0.807	0.817	0.874	0.634	2.026
		WMI2	0.855					
		WMI3	0.838					
		WMI4	0.762					
Portugal	Brand Love	BL1	0.835	0.898	0.915	0.923	0.671	3.237
n = 555		BL2_rev	*					
		BL3	0.883					
		BL4	0.901					
		BL5_rev	0.577					
		BL6	0.846					
		BL7	0.829					
	Erasmus Experience	EEX1	0.852	0.855	0.886	0.896	0.638	1.848
		EEX2	0.866					
		EEX3	0.900					
		EEX4	0.679					
		EEX5	0.663					
	E-WoM	EWM1	0.895	0.737	0.738	0.884	0.792	1.224
		EWM2	0.884					
		EWM3	*					
	Recommend	RCM1	0.890	0.904	0.906	0.940	0.839	3.047
		RCM2	0.921					
		RCM3	0.937					
	Revisit Intention	RVST1	1.000	1.000	1.000	1.000	1.000	1.852
	WoM	WM1	0.898	0.888	0.890	0.930	0.817	2.665
		WM2	0.927					
		WM3	0.885					
	WoM Intensity	WMI1	0.719	0.820	0.833	0.881	0.650	2.164
		WMI2	0.853					
		WMI3	0.827					
		WMI4	0.819					

* Item removed due to low loading.

Table 3
Assessment of Formative Measurement Model.

Dataset	Construct	Item	Outer Weights and 95 % Percentile CI	p-value	VIF	FC
Complete	Destination Image	DI1	0.317 [0.262; 0.387]	0.000	1.063	1.588
n = 2237		DI2	0.425 [0.399; 0.554]	0.000	1.714	
(All countries		DI3	0.379 [0.212; 0.372]	0.000	1.715	
combined)		DI4	0.105 [0.054; 0.191]	0.000	1.271	
		DI5	0.250 [0.173; 0.305]	0.000	1.281	
Germany	Destination Image	DI1	0.304 [0.136; 0.462]	0.000	1.049	1.509
n = 921		DI2	0.441 [0.247; 0.632]	0.000	1.690	
		DI3	0.472 [0.267; 0.648]	0.000	1.686	
		DI4	0.021 [-0.139; 0.197]	**0.726**	1.209	
		DI5	0.154 [-0.013; 0.323]	0.010	1.205	
Italy	Destination Image	DI1	0.268 [0.103; 0.438]	0.000	1.074	1.589
n = 761		DI2	0.421 [0.210; 0.628]	0.000	1.713	
		DI3	0.302 [0.098; 0.502]	0.000	1.701	
		DI4	0.140 [-0.055; 0.327]	**0.039**	1.324	
		DI5	0.319 [0.133; 0.516]	0.000	1.347	
Portugal	Destination Image	DI1	0.390 [0.178; 0.583]	0.000	1.079	1.646
n = 555		DI2	0.369 [0.098; 0.629]	0.000	1.778	
		DI3	0.327 [0.071; 0.577]	0.000	1.800	
		DI4	0.006 [-0.187; 0.204]	**0.936**	1.333	
		DI5	0.327 [0.136; 0.520]	0.000	1.358	

Note: Bold means that the indicator shows a non-significant relationship.

For the complete dataset, all items of destination image are statistically significant (with < 0.05). On the contrary, the datasets of Germany, Italy, and Portugal exhibit that DI4 is not significant. Nevertheless, this item is retained to fully capture the domain of the destination image (Lee & Lockshin, 2021; Roberts & Thatcher, 2009).

Finally, discriminant validity was evaluated using the HTMT ratio in complete and split datasets (Franke & Sarstedt, 2019; Henseler et al., 2015; Voorhees et al., 2015). Table 4 shows no discriminant validity issues for complete data and datasets from Germany, Italy, and Portugal. Indeed, all the HTMT values are lower than the threshold value of $HTMT_{.85}$ (Henseler et al., 2015). Furthermore, the interconstruct correlations between the destination image as a formative construct and other constructs are lower than 0.7, indicating sufficient discriminant validity for complete data and datasets across three countries (see Table 4).

4.2.3. Step 3: Assess measurement invariance using MICOM

A fundamental concern when comparing model estimates for substantial differences in the multigroup analysis is that the construct measures are invariant between groups (Sarstedt et al., 2011). Measurement invariance must be confirmed so that researchers can be assured that group differences in model estimates are not due to

Table 4
Assessment of Discriminant Validity using HTMT and Interconstruct Correlations.

Dataset	Construct	BL	Destination Image	E-WoM	Erasumus Experience	Recommend	Revisit Int	WoM	WoM Intensity
Complete	BL		***0.534***	0.246	0.607	0.693	0.575	0.708	0.595
n = 2237	Destination Image	0.563		***0.155***	***0.445***	***0.532***	***0.442***	***0.461***	***0.392***
(all countries combined)	E-WoM	0.292	0.177		0.290	0.254	0.148	0.314	0.362
	Erasumus Experience	0.684	0.484	0.362		0.578	0.405	0.526	0.512
	Recommend	0.767	0.564	0.305	0.658		0.524	0.710	0.601
	Revisit Int	0.609	0.442	0.170	0.441	0.555		0.480	0.429
	WoM	0.791	0.490	0.386	0.603	0.798	0.512		0.619
	WoM Intensity	0.684	0.433	0.456	0.613	0.703	0.472	0.731	
Germany	BL		***0.508***	0.171	0.596	0.672	0.540	0.697	0.560
n = 921	Destination Image	0.536		***0.102***	***0.434***	***0.519***	***0.408***	***0.451***	***0.357***
	E-WoM	0.197	0.110		0.249	0.204	0.076	0.267	0.345
	Erasumus Experience	0.675	0.473	0.304		0.565	0.377	0.519	0.502
	Recommend	0.747	0.553	0.240	0.644		0.487	0.702	0.562
	Revisit Int	0.573	0.408	0.086	0.411	0.520		0.441	0.393
	WoM	0.780	0.482	0.322	0.597	0.795	0.471		0.590
	WoM Intensity	0.645	0.395	0.426	0.605	0.662	0.433	0.699	
Italy	BL		***0.538***	0.296	0.610	0.682	0.568	0.716	0.602
n = 716	Destination Image	0.566		***0.164***	***0.461***	***0.516***	***0.444***	***0.473***	***0.379***
	E-WoM	0.354	0.193		0.302	0.277	0.188	0.342	0.379
	Erasumus Experience	0.695	0.500	0.384		0.591	0.403	0.543	0.503
	Recommend	0.754	0.547	0.333	0.678		0.497	0.715	0.622
	Revisit Int	0.600	0.444	0.218	0.437	0.524		0.494	0.437
	WoM	0.801	0.503	0.423	0.627	0.805	0.528		0.631
	WoM Intensity	0.695	0.422	0.480	0.609	0.729	0.486	0.749	
Portugal	BL		***0.571***	0.314	0.624	0.741	0.636	0.718	0.649
n = 555	Destination Image	0.602		***0.233***	***0.450***	***0.566***	***0.492***	***0.458***	***0.459***
	E-WoM	0.382	0.271		0.341	0.316	0.221	0.364	0.374
	Erasumus Experience	0.687	0.485	0.430		0.589	0.458	0.520	0.548
	Recommend	0.814	0.595	0.385	0.657		0.613	0.714	0.635
	Revisit Int	0.674	0.492	0.257	0.493	0.645		0.526	0.484
	WoM	0.797	0.484	0.452	0.588	0.792	0.557		0.651
	WoM Intensity	0.741	0.503	0.476	0.640	0.731	0.528	0.761	

Note: The HTMT result falls below the diagonal value while the above result belongs to interconstruct correlations; Values that are in bold and italic represents the result of interconstruct correlations between formative construct (i.e., destination image) and all other constructs (i.e., BL, E-WoM, Erasmus Experience, Recommend, Revisit Int, WoM, and WoM Intensity).

distinctive meanings or different interpretations of the latent variables across groups and, therefore, ensure the validity of results and conclusions (Hair et al., 2019). Conversely, if the measurement invariance is not established, then the validity of the outcomes is questionable (Hair et al., 2018).

Before performing multiple group analysis (MGA) to compare the path coefficients between the three countries' groups of Erasmus students, as proposed in this study, an invariance test using MICOM must be conducted (Henseler et al., 2016). The goal is to ascertain whether construct measurements are understood similarly across the three countries perceived by Erasmus students (Henseler et al., 2016). Three procedures should be applied to carry out the MICOM test: configural invariance, compositional invariance, and equal distribution of mean values and variances of composites.

In the initial procedure, researchers need to first establish configural invariance. Configural invariance exists when constructs are equally parameterized and estimated between groups (Henseler et al., 2016). In other words, the test ensures that each measurement model employs the same indicators across groups. This is critical in cross-cultural studies to apply good empirical research procedures to establish the equivalence of the indicators, such as back-to-back translation. However, it is not easy to determine whether the same indicators apply to all groups. Therefore, verifying whether the researcher employed the same set of indicators across the groups can be aided by qualitative investigations (Moore, Harrison, and Hair, 2021) or face and/or expert validity evaluations (see Hair et al., 2019).

Furthermore, the data treatment of the indicators and data handling must be identical or consistent across all groups (Henseler et al., 2016), mainly when processing the data of the subgroups for the configural invariance test. Moreover, according to Henseler et al. (2016), when outliers are detected, it is crucial to treat them consistently across groups. For example, multivariate outliers, such as Cook's distances, leverages, and Mahalanobis distance, can be used to detect and treat between groups (Sarstedt and Mooi, 2019). Finally, keeping an eye on straight-lining patterns is essential, which occurs when a respondent provides the same answer to almost all survey questions. These responses should be discarded during the data-cleaning stage (Hair et al., 2019). Straight-line answers reduce variability, resulting in undetected moderator (or detected but underestimated) effects in MGA (Cheah et al., 2020). In order to check for straight lining, internal consistency reliability indexes should be inspected to detect the situation. This solution was also suggested by Hair et al. (2019, p.8) that "reliability values of 0.95 and above also suggest the possibility of undesirable response patterns (e.g., straight-lining)".

Furthermore, researchers must ensure that differences in the group-specific model estimations are not caused by different algorithm settings. The choice of initial outer weights (by means value of 1.0) and the inner model weighting scheme must be predetermined when using the PLS-SEM technique. According to Hair et al. (2022a) recommendations, it is necessary to use path weighting, with a maximum of 300 iterations and a stop criterion of 10–7 in the settings of the PLS-SEM algorithm across the groups. In this study, the measurement model stage shows that the configural invariance is established across the datasets of all countries (see Tables 2 to 4).

Conducting the MICOM test requires assessing compositional invariance. Compositional invariance exists when composite scores across the groups are perfectly correlated (Henseler et al., 2016). This test can be established if the initial correlation is equal to or greater than the 5 % quantile and the p-value is non-significant. In other words, the establishment of the partial invariance can only be achieved if the permutation technique using the MICOM test shows that none of the c values are significantly different. However, when assessing the

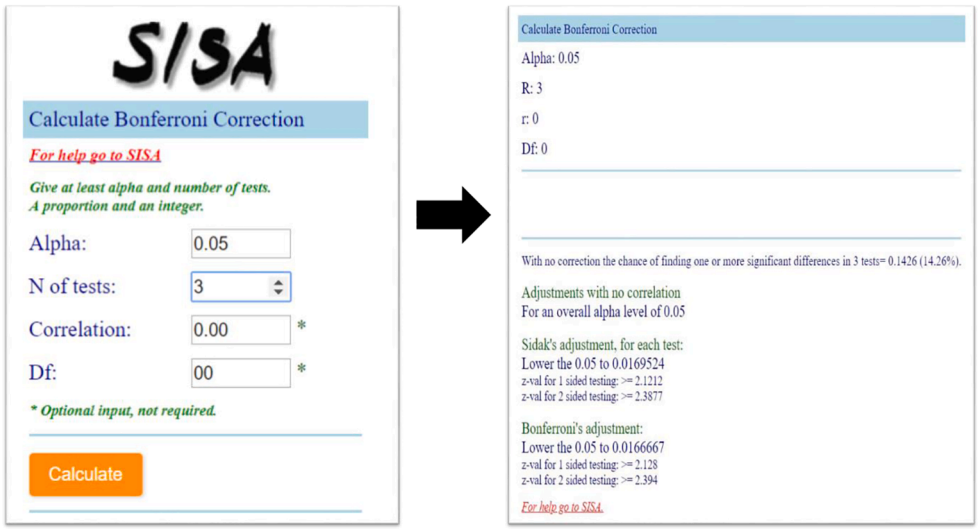


Fig. 4. Procedure for conducting the p-value of both **Šidák**'s and Bonferroni's adjustments.

invariance test of more than two groups, it is crucial to adjust the conventional p-value of 0.05 through Šidák's or Bonferroni's adjustment, to avoid the family-wise error (see Fig. 4 on the application of the p-value adjustment). If Šidák's p-value adjustment is used, researchers must adjust the significance level of 0.05 to 0.0169524 when executing the permutation test in any PLS commercial software or R statistical software.[8] In that case, the confidence intervals of the upper bound value will automatically be adjusted from 95 % to 98.31 %. Notably, the one-tailed test can be used to analyse the permutation test when directed hypotheses are involved in the study. As shown in Table 5, all permutation c value results (=1) straddle between the upper and lower bounds of the 98.31 % confidence interval, thus establishing compositional invariance in the research model for all countries' datasets.

Finally, researchers need to assess the equality of composite mean values and variances as the final requirement to establish full measurement invariance (Henseler et al., 2016). Based on Table 5, the composites' equality of mean values and variances of the composites across the three datasets do not produce all non-significance differences after applying the Šidák adjustment. The reason is that the difference in the composite's mean value and variances ratio must fall between the upper and lower bounds of the 0.84 % and 99.16 % confidence interval, especially after the p-value adjustment of Šidák's method. The MICOM procedure established partial measurement invariance, indicating a feasible comparison and interpretation of the MGA's group-specific differences in the PLS-SEM findings (Henseler et al., 2016).

4.2.4. Step 4: Determine the goal of analysis

Once measurement invariance is established (see Table 5), researchers need to determine the goal of analysis, if it is to examine the difference in a complete model or a specific path coefficient between groups. To reiterate, researchers should not evaluate group comparisons using the OTG approach (see Sarstedt et al., 2011) because this technique consistently shows significant parameter differences for many bootstrap runs, even if the pairwise evaluation of each path coefficient across the groups exhibits non-significant result. Suppose that researchers want to examine the differences in the complete model across groups because they have no established knowledge about which parameters in the model differ across groups or if they are interested in exploring whether a suggested theory or model performs differently across groups (e.g., nationality). In that case, researchers must perform the NDT to ensure differences across groups based on a complete structural model (Klesel et al., 2019; 2022). This would be similar to the ANOVA evaluation, where the F-test examines H0 and finds that all means are equal.

In contrast, if the choice is to examine the specific path coefficient across groups, researchers can immediately run the NPT for pairwise evaluation. The test randomly permutes observations between the groups and reestimates the model to derive a test statistic for the group differences (Chin and Dibbern, 2010). Drawing on the above-mentioned, this study illustrates the difference in a complete model by performing both OTG and NDT to corroborate Klesel's (2022) recommendation. Therefore, the following procedure is to assess the NPT for pairwise evaluation.

4.2.5. Step 5: Analyze and interpret the MGA test

The first step is to test for overall differences in the path coefficients of the three groups. To obtain this result, the OTG and NDT were used to assess whether the path coefficients were similar across the three samples (i.e., Germany, Italy, and Portugal).[9] The OTG analysis reveals that the null hypothesis of the eight path coefficients are similar across the five groups; hence it can be rejected in all structural models. In particular, the analysis shows F_R values of 640.74 (Destination Image → Erasmus Experience), 3426.24 (Destination Image → Destination Brand Love), 53.68 (Erasmus Experience → Destination Brand Love), 1518.23 (Destination Brand Love WOM), 14010.66 (Destination Brand Love Recommendation), 124892.99 (Destination Brand Love E-WOM), 16059.94 (Destination Brand Love → WOM Intensity), and 14501.62 (Destination Brand Love → Revisit Intention), thus rendering all differences significant at $p < 0.001$ (see Table 6). The results show significant differences between path coefficients among the three groups' samples. However, the use of the OTG produces biased results because this technique consistently exhibits significant parameter differences even though the pairwise evaluation from the NPT indicates non-significant results (see relationships of DI → EE, DI → DBL, EE → DBL, DBL →

[8] In this study, the p-value adjustment based on the Šidák procedure was used to assess the compositional invariance, the composites' equality of mean values, and variances of the composites. If SmartPLS 3 is used, the adjustment of the permutation test will be in three decimal point that is 0.017. In contrast, if SmartPLS4 is used, the adjustment of the permutation test will be in two decimal point that is 0.02.

[9] The procedure to conduct the OTG approach is in Appendix A, while the procedure to conduct NDT is in Appendix B. In order to ensure reproducibility of both results from Appendix A and Appendix B, the data can be obtained upon request from the corresponding author.

Table 5
Assessment of Measurement Invariance.

Country	Construct	Configural Invariance	Compositional Invariance		Partial Measurement Invariance Established	Equal Mean Value		Equal Variances		Full Measurement Variance Established
			c = 1	Confidence Interval		Differences	Confidence Interval	Differences	Confidence Interval	
Germany vs Italy	Brand Love	Yes	1.000	[1.000; 1.000]	Yes	−0.055	[-0.095; 0.088]	0.005	[-0.146; 0.142]	Yes
	Destination Image	Yes	0.970	[0.965; 1.000]	Yes	−0.097	[-0.095; 0.099]	−0.007	[-0.190; 0.205	Yes
	E-WoM	Yes	1.000	[0.995; 1.000]	Yes	0.033	[-0.092; 0.094]	0.072	[-0.110; 0.119]	Yes
	Erasmus Experience	Yes	1.000	[0.999; 1.000]	Yes	0.030	[-0.107; 0.099]	−0.034	[-0.146; 0.151]	Yes
	Recommend	Yes	1.000	[1.000; 1.000]	Yes	−0.138	[-0.097; 0.095]	0.010	[-0.186; 0.192]	No
	Revisit Intention	Yes	1.000	[1.000; 1.000]	Yes	−0.033	[-0.100; 0.095]	0.081	[-0.211; 0.208]	Yes
	WoM	Yes	1.000	[1.000; 1.000]	Yes	−0.099	[-0.090; 0.089]	0.022	[-0.149; 0.137]	No
	WoM Intensity	Yes	1.000	[0.999; 1.000]	Yes	−0.114	[-0.095; 0.100]	0.021	[-0.118; 0.121]	No
Germany vs Portugal	Brand Love	Yes	1.000	[1.000; 1.000]	Yes	0.017	[-0.112; 0.107]	−0.077	[-0.147; 0.142]	Yes
	Destination Image	Yes	0.976	[0.959; 1.000]	Yes	−0.071	[-0.104; 0.109]	−0.098	[-0.218; 0.218]	Yes
	E-WoM	Yes	0.997	[0.993; 1.000]	Yes	−0.072	[-0.103; 0.107]	0.062	[-0.117; 0.136]	Yes
	Erasmus Experience	Yes	0.999	[0.999; 1.000]	Yes	−0.004	[-0.103; 0.106]	−0.086	[-0.142; 0.148]	Yes
	Recommend	Yes	1.000	[1.000; 1.000]	Yes	−0.094	[-0.101; 0.109]	−0.090	[-0.199; 0.194]	Yes
	Revisit Intention	Yes	1.000	[1.000; 1.000]	Yes	0.042	[-0.102; 0.112]	−0.112	[-0.217; 0.221]	Yes
	WoM	Yes	1.000	[1.000; 1.000]	Yes	−0.069	[-0.099; 0.106]	−0.027	[-0.162; 0.152]	Yes
	WoM Intensity	Yes	1.000	[0.999; 1.000]	Yes	−0.187	[-0.102; 0.108]	−0.016	[-0.130; 0.128]	No
Italy vs Portugal	Brand Love	Yes	1.000	[1.000; 1.000]	Yes	0.070	[-0.112; 0.114]	−0.082	[-0.175; 0.168]	Yes
	Destination Image	Yes	0.985	[0.960; 1.000]	Yes	0.024	[-0.104; 0.117]	−0.115	[-0.210; 0.200]	Yes
	E-WoM	Yes	0.997	[0.995; 1.000]	Yes	−0.109	[-0.114; 0.104]	−0.004	[-0.121; 0.123]	Yes
	Erasmus Experience	Yes	0.999	[0.998; 1.000]	Yes	−0.032	[-0.111; 0.110]	−0.054	[-0.173; 0.169]	Yes
	Recommend	Yes	1.000	[1.000; 1.000]	Yes	0.042	[-0.108; 0.111]	−0.101	[-0.232; 0.221]	Yes
	Revisit Intention	Yes	1.000	[1.000; 1.000]	Yes	0.075	[-0.104; 0.107]	−0.194	[-0.224; 0.223]	Yes
	WoM	Yes	1.000	[1.000; 1.000]	Yes	0.030	[-0.114; 0.109]	−0.048	[-0.177; 0.173]	Yes
	WoM Intensity	Yes	1.000	[0.999; 1.000]	Yes	−0.074	[-0.108; 0.110]	−0.038	[-0.152; 0.135]	Yes

Note: Confidence interval is adjusted using Šidák p-value adjustment. Since SmartPLS 3 is used in the study, the adjustment of the p-value changes from 0.05 to 0.017. Therefore, the upper bound confidence interval of compositional invariance was automatically adjusted to 98.31%, while both equal mean value and variances were automatically adjusted to a lower bound of 0.84% and an upper bound of 99.16%.

WOM, and DBL → Recommend).

In contrast, the MGA assessment was extended using the NDT suggested by Klesel et al. (2019). Table 7 shows a significant result for dL but a non-significant result for dG. Thus, the findings indicate that comparing the complete model between the three countries is necessary. This result confirms Klesel et al.'s (2019) work that the dG test is stricter than the dL test when detecting heterogeneity (or differences) in the structural model across groups. In other words, researchers should not be too rigid when using the NDT approach because the result could achieve significant differences in a specific part of the model between different groups. Hence, NDT would produce a non-significant result on either dG or dL. On the other hand, if both dG or dL show a non-significant result, the group comparisons should be rejected, or the grouping variable should be respecified. Additionally, researchers could also check for data and measurement problems.

Since both OTG and NDT do not provide clear parameter information on whether there are specific differences between path coefficients between group comparisons, pairwise comparisons are performed (Hair et al., 2018). The permutation test for multigroup comparisons (Sarstedt et al., 2011) was used to conduct pairwise comparisons. If there is a difference in the path coefficients between the groups, researchers can continue to evaluate and report the structural model for each group separately. If there is no difference in the path coefficients between the two groups, it is recommended to check for data and measurement problems.

To assess the pairwise comparison of more than two groups – in the

Table 6
Assessment of the Omnibus Test of Group (OTG) and Pairwise Comparison using the Non-Parametric Permutation-Based Test (NPT).

Relationship	Comparison	diff	Permutation Test		
			Conventional p-value: 0.05	Šidák's p-value adjustment: 0.0169524	Bonferroni's p-value adjustment: 0.0166667
DI → EE	Germany vs Italy	−0.021	0.630	0.637	0.610
[Fr: 640.74; p-value: <0.000]	Germany vs Portugal	−0.018	0.695	0.731	0.702
	Italy vs Portugal	0.004	0.956	0.953	0.958
DI → DBL	Germany vs Italy	−0.027	0.542	0.566	0.573
[Fr: 3426.24; p-value: <0.000]	Germany vs Portugal	−0.053	0.286	0.278	0.281
	Italy vs Portugal	−0.025	0.601	0.593	0.600
EE → DBL	Germany vs Italy	0.005	0.898	0.908	0.907
[Fr: 53.68; p-value: <0.000]	Germany vs Portugal	0.002	0.975	0.979	0.977
	Italy vs Portugal	−0.004	0.934	0.954	0.932
DBL → WOM	Germany vs Italy	−0.019	0.456	0.468	0.488
[Fr: 1518.23; p-value: <0.000]	Germany vs Portugal	−0.021	0.496	0.498	0.494
	Italy vs Portugal	−0.002	0.959	0.965	0.974
DBL → Recommend	Germany vs Italy	−0.010	0.760	0.773	0.776
[Fr: 14010.66; p-value: <0.000]	Germany vs Portugal	−0.068	**0.023**	0.031	0.035
	Italy vs Portugal	−0.059	**0.033**	0.068	0.076
DBL → EWOM	Germany vs Italy	−0.125	**0.002**	**0.004**	**0.005**
[Fr: 24892.99; p-value: <0.000]	Germany vs Portugal	−0.143	**0.001**	**0.003**	**0.004**
	Italy vs Portugal	−0.018	0.734	0.745	0.720
DBL → WOM Intensity	Germany vs Italy	−0.042	0.228	0.237	0.220
[Fr: 16059.94; p-value: <0.000]	Germany vs Portugal	−0.088	**0.004**	**0.009**	**0.008**
	Italy vs Portugal	−0.046	0.208	0.223	0.212
DBL → RI	Germany vs Italy	−0.028	0.516	0.503	0.494
[Fr: 14501.62; p-value:<0.000]	Germany vs Portugal	−0.096	**0.005**	**0.009**	**0.009**
	Italy vs Portugal	−0.068	0.149	0.068	0.112

Note: Bold values indicate statistically significant.

Table 7
NDT approach by Klesel et al. (2019).

Ho: Model-Implied Indicator Covariance Matrix is Equal Across Groups			
Distance Measure	Test Statistic	p-value	Decision
d_G	0.144	0.581	Do not reject
d_L	2.074	0.038	Reject

case of this study, Germany and Italy, Germany and Portugal, and Italy and Portugal – the Bonferroni or the Šidák correction is compulsory to avoid family-wise error. Researchers can use the Bonferroni-Šidák Correction calculator[10] if they encounter difficulties calculating the Bonferroni or the Šidák correction (see Fig. 4). The calculator requires researchers to set an alpha value and the number of tests (which depends on the number of groups to be compared, which in this study is 3). Regarding the correlation and degree of freedom (Df), researchers can use the default value of 0. Considering an alpha value of 0.05, the p-value of Šidák's adjustment is 0.0169524 and the p-value of Bonferroni's adjustment is 0.0166667. The differences in path coefficient in three pairs of comparisons using the conventional p-value (<0.05), the p-value of Šidák's adjustment (<0.0169524), and the p-value of Bonferroni's adjustment (<0.0166667) when using a two-tailed test are shown in Table 7.

Specifically, the conventional p-value result suggests significant differences regarding the relationship of destination brand love on the outcome of recommendation, E-WOM, WOM Intensity, and Revisit Intention. First, only the comparisons between Germany and Portugal (|diff| = −0.068; p-value = 0.023) and Italy and Portugal (|diff| = −0.059; p-value = 0.033) show a significant relationship between destination brand love and recommendation (see Table 6). However, the result of the paired comparison of these countries is not significant when using the adjustment of the p-value by both Šidák and Bonferroni.

When the conventional p-value is used, the German sample significantly differs from the Italian sample (|diff| = −0.125; p-value = 0.002) and the Portuguese sample (|diff| = 0.143; p-value = 0.001), particularly for the relationship between destination brand love and E-WOM. Similarly, the pair-comparison result for these countries is also significant when using Šidák's and Bonferroni's adjustments for this relationship with a p-value ≤ 0.005.

On the other hand, the effect of the destination brand love on WOM Intensity is also significant when using the conventional p-value, Šidák's adjustment, and Bonferroni's adjustment. In particular, there are path differences between samples from Germany and Portugal (|diff| = −0.088; p-value = 0.004, 0.009, and 0.008). Finally, the conventional p-value, Šidák's adjustment, and Bonferroni's adjustment results of the relationship between the destination brand love and revisit intention show that the Germany sample differs significantly from the Portugal sample (|diff| = −0.096; p-value = 0.005, 0.009; 0.009).

In conclusion, the pair-comparison analysis shows that some results are not significant when using both Šidák's and Bonferroni's p-value adjustment. However, both adjustments correspond to the same result. Therefore, once the results of the pair-comparison show a significant difference, further interpretation of the path coefficient of a specific relationship between groups is crucial. This can be assessed by comparing the strength of each value of the path coefficient or the quality criteria (coefficient of determination, effect size, and predictive

[10] The Bonferroni-Šidák Correction calculator link: https://www.quantitativeskills.com/sisa/calculations/bonfer.htm.

Table 8
Assessment of the Structural Model.

Data Set	Relationship	Std Beta	Std Error	*t*-value	P Values	Percentile 95 % CI LB	Percentile 95 % CI UB	VIF	f^2	R^2	Q^2_predict
Complete	Destination Image → Erasmus Experience	0.445	0.020	22.670	0.000	0.414	0.478	1.000	NA	0.198	0.195
(All countries combined)	Destination Image → DBL	0.329	0.018	18.088	0.000	0.301	0.360	1.247	0.160	0.455	0.282
	Erasmus Experience → DBL	0.460	0.019	24.689	0.000	0.428	0.488	1.247	0.311		
	DBL → E-WoM	0.246	0.020	12.274	0.000	0.213	0.280	1.000	NA	0.061	0.023
	DBL → Recommend	0.693	0.013	54.289	0.000	0.671	0.714	1.000	NA	0.480	0.255
	DBL → Revisit Int	0.575	0.017	33.853	0.000	0.544	0.603	1.000	NA	0.331	0.176
	DBL → WoM	0.708	0.012	58.146	0.000	0.691	0.729	1.000	NA	0.502	0.204
	DBL → WoM Intensity	0.595	0.014	43.567	0.000	0.574	0.618	1.000	NA	0.354	0.147
Germany	Destination Image → Erasmus Experience	0.438	0.028	15.393	0.000	0.357	0.504	1.000	NA	0.192	0.184
	Destination Image → DBL	0.310	0.031	9.928	0.000	0.229	0.385	1.238	0.137	0.433	0.253
	Erasmus Experience → DBL	0.460	0.030	15.549	0.000	0.384	0.534	1.238	0.302		
	DBL → E-WoM	0.171	0.032	5.367	0.000	0.080	0.219	1.000	NA	0.029	0.009
	DBL → Recommend	0.672	0.021	31.766	0.000	0.612	0.706	1.000	NA	0.452	0.227
	DBL → Revisit Int	0.540	0.029	18.860	0.000	0.462	0.586	1.000	NA	0.292	0.147
	DBL → WoM	0.697	0.019	37.066	0.000	0.646	0.725	1.000	NA	0.486	0.184
	DBL → WoM Intensity	0.560	0.023	24.530	0.000	0.496	0.596	1.000	NA	0.314	0.115
Italy	Destination Image → Erasmus Experience	0.460	0.036	12.917	0.000	0.395	0.513	1.000	NA	0.211	0.200
	Destination Image → DBL	0.338	0.034	10.028	0.000	0.281	0.391	1.268	0.167	0.462	0.289
	Erasmus Experience → DBL	0.455	0.035	13.029	0.000	0.396	0.513	1.268	0.303		
	DBL → E-WoM	0.296	0.034	8.825	0.000	0.239	0.351	1.000	NA	0.088	0.025
	DBL → Recommend	0.682	0.026	26.589	0.000	0.637	0.722	1.000	NA	0.465	0.247
	DBL → Revisit Int	0.568	0.032	17.677	0.000	0.514	0.620	1.000	NA	0.323	0.175
	DBL → WoM	0.716	0.021	33.947	0.000	0.680	0.748	1.000	NA	0.513	0.218
	DBL → WoM Intensity	0.602	0.026	23.541	0.000	0.557	0.641	1.000	NA	0.363	0.141
Portugal	Destination Image → Erasmus Experience	0.456	0.040	11.443	0.000	0.381	0.511	1.000	NA	0.208	0.193
	Destination Image → DBL	0.363	0.034	10.834	0.000	0.303	0.415	1.263	0.207	0.494	0.314
	Erasmus Experience → DBL	0.459	0.034	13.482	0.000	0.402	0.513	1.263	0.329		
	DBL → E-WoM	0.314	0.039	7.997	0.000	0.247	0.374	1.000	NA	0.099	0.052
	DBL → Recommend	0.741	0.020	36.683	0.000	0.705	0.772	1.000	NA	0.548	0.292
	DBL → Revisit Int	0.636	0.028	22.545	0.000	0.587	0.678	1.000	NA	0.405	0.213
	DBL → WoM	0.718	0.024	30.380	0.000	0.675	0.754	1.000	NA	0.516	0.205
	DBL → WoM Intensity	0.649	0.025	25.721	0.000	0.602	0.688	1.000	NA	0.421	0.201

power) in the evaluation of the structural model (see Table 8).

Before evaluating structural models, it is critical to check that collinearity issues are not linked to assessing the structural model. Therefore, full collinearity variance inflation factors (VIFs) are evaluated as a viable option to assist in identifying multicollinearity issues. The results of the full collinearity test are shown in Table 8. Each construct's VIF score is below the threshold value of 3.3, demonstrating that collinearity is not an issue (Hair et al. 2022a).

The causal relationships between the model's constructs are described by the structural model (i.e., path coefficients and the coefficient of determination, R^2 value) (Hair et al. 2022a). The bootstrapping technique with 10,000 samples is used to assess the significance of the path coefficient (Becker et al., 2022; Hair et al. 2022a Streukens & Leroi-Werelds, 2016). Firstly, the path coefficients are positively significant for the Complete, Germany, Italy, and Portugal datasets (see Table 8). Notwithstanding the significant relationships, it is critical to examine the effect sizes of the paths' (f^2) (Cohen, 1988). Based on Table 8, the complete data ($f^2 = 0.160$), Italy data ($f^2 = 0.167$), and Portugal data ($f^2 = 0.207$) show a moderate effect on the relationship between Destination Image and Destination Brand Love, except for the German data ($f^2 = 0.137$) that exhibits small effect size. On the other hand, the Complete data ($f^2 = 0.311$), German data ($f^2 = 0.302$), and Italy data ($f^2 = 0.303$) show a moderate effect on the relationship between Erasmus Experience and Destination Brand Love, except for Portugal data ($f^2 = 0.329$) that exhibits large effect size.

Table 8 also shows the R^2 values for each sample to illustrate the variance explained. The 19 % and above values indicate that all data have the adequate capacity to explain the Destination Image in the Erasmus Experience during their study in Europe. On the other hand, the relationship between the Destination Image and the Erasmus Experience in Destination Brand Love, where R^2 indicates 40 % and above. Finally, the results (Recommendation, Revisit Intention, WOM, and WOM intensity) of Destination Brand Love show an adequate explanatory power of more than 20 % and above, compared to e-WOM, which exhibits<10 %. Finally, since PLS-SEM prioritizes causal prediction results over theory testing, the PLSpredict technique was used to understand the predictive relevance of the proposed path model (Shmueli et al., 2019). Endogenous latent variables show a value greater than zero, indicating that all data sets have acceptable predictive accuracy (see Table 8).

5. Discussion and recommendations

A close examination of MGA in PLS-SEM research demonstrates the lack of rigor in applying this methodology. The application of MGA in PLS-SEM requires specific details so that the data can be adequately analyzed. Researchers in business and management fields often neglect to assess measurement invariance, making their conclusions about model relationships questionable. However, MGA in PLS-SEM studies rarely exploit the potential to examine three or more groups and therefore miss opportunities to conduct more meaningful research. Several vital issues concerning MGA with three or more groups have not yet been well investigated, such as handling family-wise errors. Therefore, this research makes an essential contribution to this development. This study addresses the most pressing questions of researchers about using PLS-SEM, that is, how to fully utilize the MGA technique with three or more groups, which enables a researcher to apply it to business

Table 9
Overview.

No.	Question	Recommendations
1	Before conducting MGA with more than two groups, what are the data preparations and considerations?	• Prior knowledge of theoretically plausible models is necessary to ensure a meaningful group comparison study. When detecting and defining groups, the FIMIX-PLS procedure or cluster analysis should be considered.Converting a continuous moderator to a categorical variable of more than two groups (low, medium, and high conditions) should be avoided when conducting MGA because it can reduce the statistical power to detect path differences. However, when a dichotomization is needed from a continuous variable, researchers should consider the use of dividing points (i.e., percentiles) . In this case, researchers should present strong underlying reasoning and justification.Planning well research design is critical, particularly in collecting the right amount of data when assessing the heterogeneity effects of more than two groups. In other words, the sample size must be appropriate and representative of the population based on the study's research objective (Moore, Harrison & Hair, 2021). Online surveys, such as Amazon Mechanical Turk (MTurk), can be used to collect a large sample size of subsample groups. This ensures that sufficient observations are collected when conducting a study with more than two groups. Additionally, researchers should try to collect data so that each subgroup has a comparable size to avoid underestimating moderating effects (see Aguinis et al., 2017). Finally, adequate statistical power should be achieved (Hair et al., 2022ba). The gamma exponential method and the inverse square root method or Hair et al.'s (2022a) power analysis table can be used to determine the minimum sample size requirement.
2	What measurement tests should be considered before assessing the MGA with more than two groups?	• The assessment of measurement models – whether reflective or formative - must meet the satisfactory criteria or rule of thumb (see Hair et al. 2022a). Researchers could depend on the confirmatory tetrad analysis (CTA-PLS) (see Gudergan et al., 2008) to avoid any potential model misspecification when assessing the measurement model.
3	Should the comparison of more than two groups be analyzed with or without measurement invariance?	• Evaluation of measurement invariance is essential, particularly in using the measurement invariance of composite models (MICOM) in PLS-SEM (see Henseler et al., 2016). If partial or full invariance is established, researchers could begin to evaluate the MGA with more than two groups to ensure that the differences between groups in the model estimates are not the result of, for example, group-specific response styles (Henseler et al., 2016).When assessing the measurement invariance test using MICOM, the adjustment of the conventional p-value (i.e., Šidák's and/or Bonferroni's) is needed to ensure that no family-wise errors occur in the study.
4	Should group comparisons of more than two groups be analyzed using Omnibus Test of Group (OTG) differences?	• The OTG approach (Sarstedt et al., 2011) should be avoided because this procedure consistently demonstrates significant parameter differences for many bootstrap runs, even if the pairwise evaluation of each path coefficient across the groups exhibits non-significant results. In other words, the OTG almost always rejects the null hypothesis of no group differences. Therefore, it is not recommendable (Klesel et al., 2022).
5	When should researchers use the Non-Parametric Distance Based Test (NDT) or the non-parametric permutation-based test (NPT)?	• If the aim is to examine the difference in the complete model across groups, the non-Parametric Distance Based Test (NDT) can be performed to ensure the differences across groups based on a complete structural model (Klesel et al., 2019; 2022).If the objective is to examine the specific path differences across groups, researchers can subsequently perform the NPT to achieve the pairwise comparison (Chin & Dibbern, 2010).
6	How should Non-Parametric Distance Based Test (NDT) be assessed?	• If either or both criteria (dG or dL) in NDT show a significant result (rejecting H_0), researchers should perform the pairwise evaluation of each path coefficient across the groups (which can be considered as the post-hoc tests) using the non-parametric permutation-based test (NPT) (Chin 2003; Chin and Dibbern, 2010) with the Šidák's and/or Bonferroni's p-value adjustment to manage the family-wise error (see Hair et al., 2018).However, if the NDT exhibits a non-significant result, the group comparisons should be rejected, or the grouping variable should be respecified based on a well-established theoretical justification (see Hair et al. 2022a) . The findings of the present study are consistent with the Klesel et al.'s (2019) research that the dG criterion is stricter in producing higher rejection rates of H_0 than the dL criterion. However, Klesel et al. (2019) also highlighted that both criteria (dG and dL) in NDT could perform slightly worse – by means of being unable to detect heterogeneity – when the data are non-normally distributed and if the path differences are minor across groups (approximate β difference of 0.1). Therefore, following Klesel et al.'s (2019) recommendation, a moderate structural difference (approximate β difference of 0.2 across groups) requires 200 observations per group (or more) to detect heterogeneity using the dG criterion, mainly when data are normally distributed. In addition, for non-normal data, 300 observations per

(*continued on next page*)

Table 9 (*continued*)

No.	Question	Recommendations
		group (or more) with a moderate structural difference are necessary to detect heterogeneity for both criteria of dG and dL. To conclude, when using dG or dL in NDT, researchers should consider the normality of the data, sample size, and size of the differences in the path coefficient in the structural model.
7	How should the non-parametric permutation-based test (NPT) be assessed?	• To avoid family-wise error, either Šidák's or Bonferroni's p-value adjustment must be used to assess the non-parametric permutation-based test (NPT) (Chin 2003; Chin and Dibbern, 2010). A one-tailed or two-tailed test can be applied for this pairwise group comparison test. However, it depends on how the research hypotheses are proposed through either directional or non-directional research hypotheses.If the pairwise group comparison shows a significant result, there are group differences regarding the specific comparison of path coefficients (direct or indirect effects) . If the NPT does not show any significant specific comparison of path coefficients, researchers can check on their data and measurement issues.
8	Should the structural model be further assessed?	• If neither NDT nor NPT show a promising result, it is worth investigating whether there are group differences in the specific comparison of path coefficients (i.e., direct or indirect effects; see Cheah et al., 2021) as well as other quality criteria (such as coefficient of determination, by means R^2, effect size (f^2), and PLS predict, model selection criteria, and cross-validated predictive ability test (CVPAT) (see Liengaard et al., 2021; Sarstedt et al., 2022a; Sharma et al., 2022).

research in general. Table 9 summarizes the key questions and the proposed answers and recommendations, which will be valuable for researchers to apply the method effectively.

6. Conclusions

In summary, this study contributes in several ways to the body of knowledge about PLS-MGA. First, it evidences that most PLS-MGA research only compares two groups. Second, it reveals that much research does not assess measurement invariance. Third, it provides an empirical example that compares more than two groups, using an appropriate technique such as the NDT, highlighting the importance of handling family error when comparing more than two groups (Šidák's and Bonferroni's adjustment of the p-value). Indeed, adjusting the conventional p-value with Šidák's or Bonferroni's adjustment is essential, as some paths were not significant after adjusting the conventional p-value. Fourth, this study may be the first to present a detailed empirical example of PLS-MGA with more than two groups, providing clear guidelines that can motivate future PLS-MGA research with more than two groups. Finally, this study encourages PLS-SEM software developers to implement NDT to allow researchers to apply the MGA of more than two groups without using R programming.

7. Future studies

Future research should address new and expanded issues on the topics discussed in this paper. For instance, researchers could benchmark the systematic review and compare the MGA of more than two groups, particularly between PLS-SEM and covariance-based SEM (Hair, Black, Babin & Anderson, 2019) or consistent PLS-SEM algorithm (Dijkstra & Henseler, 2015a, 2015b). Second, future studies could also use the MGA criteria (i.e., measurement invariance using MICOM, OTG, NDT, Šidák, and Bonferroni) to examine other composite-based SEM methods such as generalized structured component analysis (Hwang et al., 2010), and regularized canonical correlation analysis (Tenenhaus & Tenenhaus, 2011). Finally, future studies could also perform Monte Carlo simulation to discover which p-value adjustment test (i.e., Bonferroni, Šidák, Holm, and other approaches) is more robust when dealing with multiple pairwise comparisons of more than two groups.

CRediT authorship contribution statement

Jun-Hwa Cheah: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **Suzanne Amaro:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **José L. Roldán:** Writing – review & editing, Validation, Supervision, Formal analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is funded by National Funds through the FCT – Foundation for Science and Technology, I.P., within the scope of the project Ref. UIDB/05583/2020. Furthermore, we would like to thank the Research Centre in Digital Services (CISeD) and the Instituto Politécnico de Viseu for their support.

Appendix A:. Steps to execute OTG by Sarstedt et al. (2011)

```
rm(list = ls())                 #deletes all existing objects
set.seed(5)                     #arbitrary random number generator seed
setwd("C:/OTG")                 #set working directory, which includes all relevant files
#contains the bootstrap results of one group in each column
delimiter <- ","                #define delimiter
separator <- "."                #define decimal sepator
MCruns <- 500                   #define number of Monte Carlo runs
```

(*continued on next page*)

(*continued*)

```
rm(list = ls())                #deletes all existing objects
#OTG for the Destination Image->Erasmus Experience relationship
data_file_name <- "DI-EE.csv"    #select data file for one parameter
FR <- 0*(1:MCruns)
data <- as.matrix(read.table(file = data_file_name, sep = delimiter, dec = separator))
n < -dim(data)[1]
F0 <- n * var(colMeans(data)) / mean(diag(var(data)))
for (i in 1:MCruns){
mcdata <- t(apply(data, 1, sample))
FR[i] <- n * var(colMeans(mcdata)) / mean(diag(var(mcdata)))
}
P <- sum(FR>=F0) / MCruns
P
#OTG for the Erasmus Experience->Destination Brand Love relationship
data_file_name <- "EE-DBL.csv"    #select data file for one parameter
FR <- 0*(1:MCruns)
data <- as.matrix(read.table(file = data_file_name, sep = delimiter, dec = separator))
n < -dim(data)[1]
F0 <- n * var(colMeans(data)) / mean(diag(var(data)))
for (i in 1:MCruns){
mcdata <- t(apply(data, 1, sample))
FR[i] <- n * var(colMeans(mcdata)) / mean(diag(var(mcdata)))
}
P <- sum(FR>=F0) / MCruns
P
#OTG for the Destination Image->Destination Brand Love relationship
data_file_name <- "DI-DBL.csv"    #select data file for one parameter
FR <- 0*(1:MCruns)
data <- as.matrix(read.table(file = data_file_name, sep = delimiter, dec = separator))
n < -dim(data)[1]
F0 <- n * var(colMeans(data)) / mean(diag(var(data)))
for (i in 1:MCruns){
mcdata <- t(apply(data, 1, sample))
FR[i] <- n * var(colMeans(mcdata)) / mean(diag(var(mcdata)))
}
P <- sum(FR>=F0) / MCruns
P
#OTG for the Destination Brand Love->Word of Mouth relationship
data_file_name <- "DBL-WOM.csv"    #select data file for one parameter
FR <- 0*(1:MCruns)
data <- as.matrix(read.table(file = data_file_name, sep = delimiter, dec = separator))
n < -dim(data)[1]
F0 <- n * var(colMeans(data)) / mean(diag(var(data)))
for (i in 1:MCruns){
mcdata <- t(apply(data, 1, sample))
FR[i] <- n * var(colMeans(mcdata)) / mean(diag(var(mcdata)))
}
P <- sum(FR>=F0) / MCruns
P
#OTG for the Destination Brand Love->Recommendation relationship
data_file_name <- "DBL-RECOM.csv"    #select data file for one parameter
FR <- 0*(1:MCruns)
data <- as.matrix(read.table(file = data_file_name, sep = delimiter, dec = separator))
n < -dim(data)[1]
F0 <- n * var(colMeans(data)) / mean(diag(var(data)))
for (i in 1:MCruns){
mcdata <- t(apply(data, 1, sample))
FR[i] <- n * var(colMeans(mcdata)) / mean(diag(var(mcdata)))
}
P <- sum(FR>=F0) / MCruns
P
#OTG for the Destination Brand Love->Electronic Word of Mouth relationship
data_file_name <- "DBL-EWOM.csv"    #select data file for one parameter
FR <- 0*(1:MCruns)
data <- as.matrix(read.table(file = data_file_name, sep = delimiter, dec = separator))
n < -dim(data)[1]
F0 <- n * var(colMeans(data)) / mean(diag(var(data)))
for (i in 1:MCruns){
mcdata <- t(apply(data, 1, sample))
FR[i] <- n * var(colMeans(mcdata)) / mean(diag(var(mcdata)))
}
P <- sum(FR>=F0) / MCruns
P
#OTG for the Destination Brand Love->Word of Mouth Intensity relationship
data_file_name <- "DBL-EWOMINT.csv"    #select data file for one parameter
FR <- 0*(1:MCruns)
data <- as.matrix(read.table(file = data_file_name, sep = delimiter, dec = separator))
```

(*continued on next page*)

(*continued*)

```
rm(list = ls())                #deletes all existing objects
n < -dim(data)[1]
F0 <- n * var(colMeans(data)) / mean(diag(var(data)))
for (i in 1:MCruns){
mcdata <- t(apply(data, 1, sample))
FR[i] <- n * var(colMeans(mcdata)) / mean(diag(var(mcdata)))
}
P <- sum(FR>=F0) / MCruns
P
#OTG for the Destination Brand Love->Revisit Intention relationship
data_file_name <- "DBL-RI.csv"    #select data file for one parameter
FR <- 0*(1:MCruns)
data <- as.matrix(read.table(file = data_file_name, sep = delimiter, dec = separator))
n < -dim(data)[1]
F0 <- n * var(colMeans(data)) / mean(diag(var(data)))
for (i in 1:MCruns){
mcdata <- t(apply(data, 1, sample))
FR[i] <- n * var(colMeans(mcdata)) / mean(diag(var(mcdata)))
}
P <- sum(FR>=F0) / MCruns
P
```

Appendix B:. Steps to execute NDT by Klesel et al. (2019)

```
# Import Data
library(readxl)
ErasmusPLS <- read_excel("C:/Users/User/Desktop/ErasmusPLS.xlsx")
View(ErasmusPLS)
library(cSEM)
model<-"
# Structural Model
EEX ~ DI
DBL ~ DI + EEX
WOM ~ DBL
RCM ~ DBL
EWM ~ DBL
WMI ~ DBL
RVST ~ DBL
# Composite model
EEX <~ EEX1 + EEX2 + EEX3 + EEX4 + EEX5
DI <~ DI1 + DI2 + DI3 + DI4 + DI5
DBL <~ BL1 + BL3 + BL4 + BL5_rev + BL6 + BL7
WOM <~ WM1 + WM2 + WM3
RCM <~ RCM1 + RCM2 + RCM3
EWM <~ EWM1 + EWM2
WMI <~ WMI1 + WMI2 + WMI3 + WMI4
RVST <~ RVST1
"
# Perform estimation
res_pls <- csem(.data = ErasmusPLS,.model = model)
# Get summary
summarize(res_pls)
as.factor(ErasmusPLS$Country)
# Solution 1: You can use the.id argument and the original dataset
out <- csem(ErasmusPLS, model,.resample_method = "bootstrap",.R = 1000,.id='Country')
summarize(out)
#Solution 2: You can create a list of datasets and provide this list to the csem function
dat < -list(dat1 = ErasmusPLS[ErasmusPLS$Country==1,], dat2 = ErasmusPLS[ErasmusPLS$Country==2,], dat3 =
  ErasmusPLS[ErasmusPLS$Country==3,])
# Please have a look at the help file of the testMGD function, it has various arguments
# that you can use but in our case we focused on NDT by Klesel et al. (2019)
outMGD = testMGD(out,.R_permutation = 1000,.approach_mgd = 'Klesel')
outMGD
```

References

Aguinis, H., Edwards, J. R., & Bradley, K. J. (2017). Improving our understanding of moderation and mediation in strategic management research. *Organizational Research Methods, 20*(4), 665–685. https://doi.org/10.1177/1094428115627498

Aguirre-Urreta, M., & Rönkkö, M. (2015). Sample Size Determination and Statistical Power Analysis in PLS Using R: An Annotated Tutorial. *Communications of the Association for Information Systems, 36*, 33–51. https://doi.org/10.17705/1CAIS.03603

Ahmad, W., Kim, W. G., Choi, H. M., & Haq, J. U. (2021). Modeling behavioral intention to use travel reservation apps: A cross-cultural examination between US and China. *Journal of Retailing and Consumer Services, 63*, Article 102689. https://doi.org/10.1016/j.jretconser.2021.102689

Amaro, S., Barroco, C., & Antunes, J. (2020). Exploring the antecedents and outcomes of destination brand love. *Journal of Product & Brand Management, 30*(3), 433–448. https://doi.org/10.1108/JPBM-08-2019-2487

Arsenovic, J., De Keyser, A., Edvardsson, B., Tronvoll, B., & Gruber, T. (2021). Justice (is not the same) for all: The role of relationship activity for post-recovery outcomes. *Journal of Business Research, 134*, 342–351. https://doi.org/10.1016/j.jbusres.2021.05.031

Assaker, G., Hallak, R., Assaf, A. G., & Assad, T. (2015). Validating a Structural Model of Destination Image, Satisfaction, and Loyalty Across Gender and Age: Multigroup Analysis with PLS-SEM. *Tourism Analysis, 20*(6), 577–591. https://doi.org/10.3727/108354215X14464845877797

Basco, R., Hernández-Perlines, F., & Rodríguez-García, M. (2020). The effect of entrepreneurial orientation on firm performance: A multigroup analysis comparing China, Mexico, and Spain. *Journal of Business Research, 113*, 409–421. https://doi.org/10.1016/j.jbusres.2019.09.020

Becker, J.-M., Cheah, J.-H., Gholamzade, R., Ringle, C. M., & Sarstedt, M. (2022). PLS-SEM's most wanted guidance. *International Journal of Contemporary Hospitality Management.*

Bortz, J., Lienert, G. A., & Boehnke, K. (2003). *Verteilungsfreie Methoden in der Biostatistik* ((3 ed.).). Springer.

Bruhn, M., Georgi, D., & Hadwich, K. (2008). Customer equity management as formative second-order construct. *Journal of Business Research, 61*(12), 1292–1301. https://doi.org/10.1016/j.jbusres.2008.01.016

Carranza, R., Díaz, E., Martín-Consuegra, D., & Fernández-Ferrín, P. (2020). PLS–SEM in business promotion strategies. A multigroup analysis of mobile coupon users using MICOM. *Industrial Management & Data Systems, 120*(12), 2349–2374. https://doi.org/10.1108/IMDS-12-2019-0726

Cheah, J. H., Thurasamy, R., Memon, M. A., Chuah, F., & Ting, H. (2020). Multigroup analysis using SmartPLS: Step-by-step guidelines for business research. *Asian. Journal of Business Research, 10*(3).

Cheah, J. H., Nitzl, C., Roldán, J. L., Cepeda-Carrion, G., & Gudergan, S. P. (2021). A primer on the conditional mediation analysis in PLS-SEM. *ACM SIGMIS Database: The DATABASE for Advances. Information Systems, 52*(SI), 43–100. https://doi.org/10.1145/3505639.3505645

Cheah, J. H., Sarstedt, M., Ringle, C. M., Ramayah, T., & Ting, H. (2018). Convergent validity assessment of formatively measured constructs in PLS-SEM: On using single-item versus multi-item measures in redundancy analyses. *International Journal of Contemporary Hospitality Management, 30*(11), 3192–3210. https://doi.org/10.1108/IJCHM-10-2017-0649

Chin, W. W. (2003). A permutation procedure for multigroup comparison of PLS models. In M. Vilares, M. Tenenhaus, P. Coelho, V. Esposito Vinzi, & A. Morineau (Eds.), *Proceedings of the International Symposium PLS'03*. PLS and related methods (pp. 33–43).

Chin, W., Cheah, J.-H., Liu, Y., Ting, H., Lim, X.-J., & Cham, T. H. (2020). Demystifying the role of causal-predictive modeling using partial least squares structural equation modeling in information systems research. *Industrial Management & Data Systems, 120*(12), 2161–2209. https://doi.org/10.1108/IMDS-10-2019-0529

Chin, W. W., & Dibbern, J. (2010). An introduction to a permutation-based procedure for multigroup PLS analysis: Results of tests of differences on simulated data and a cross cultural analysis of the sourcing of information system services between Germany and the USA. In V. Esposito Vinzi, W. W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of partial least squares* (pp. 171–193). Springer.

Chin, W. W., Mills, A. M., Steel, D. J., & Schwarz, A. (2012). Multi-Group Invariance Testing: An Illustrative Comparison of PLS Permutation and Covariance-Based SEM Invariance Analysis. In 7th International Conference on Partial Least Squares and Related Methods (p. 11). Houston, Texas, USA.

Cohen, J., (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Lawrence Erlbaum.

Franke, G., & Sarstedt, M. (2019). Heuristics versus statistics in discriminant validity testing: A comparison of four procedures. *Internet Research, 29*(3), 430–447. https://doi.org/10.1108/IntR-12-2017-0515

Ghasemy, M., Mohajer, L., Cepeda-Carrión, G., & Roldán, J. L. (2020). Job performance as a mediator between affective states and job satisfaction: A multigroup analysis based on gender in an academic environment. *Current Psychology.*. https://doi.org/10.1007/s12144-020-00649-9

Gudergan, S. P., Ringle, C. M., Wende, S., & Will, A. (2008). Confirmatory tetrad analysis in PLS path modeling. *Journal of Business Research, 61*(12), 1238–1249. https://doi.org/10.1016/j.jbusres.2008.01.012

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* ((8th ed.).). Cengage Learning.

Hair, J. F., Howard, M., & Nitzl, C. (2020). Assessing measurement model quality in PLS-SEM using confirmatory composite analysis. *Journal of Business Research, 109*, 101–110.

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2022b). *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R.* Springer.

Hair, J. F., Page, M., & Brunsveld, N. (2020). *Essentials of Business Research Methods* ((4th ed.).). Routledge.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Hair, J. F., & Sarstedt, M. (2019). Factors versus Composites: Guidelines for Choosing the Right Structural Equation Modeling Method. *Project Management Journal, 50*(6), 619–624. https://doi.org/10.1177/8756972819882132

Hair, J. F., Sarstedt, M., Ringle, C. M., & Gudergan, S. P. (2018). Advanced issues in partial least squares structural equation modeling. *Sage*. Publications.

Hair, J. F. J., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022a). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* ((3rd ed.).). SAGE Publications.

Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology, 67*(3), 451–470. https://doi.org/10.1111/bmsp.12028

Henseler, J. (2012). PLS-MGA: A non-parametric approach to partial least squares-based multigroup analysis. In W. A. Gaul, A. Geyer-Schulz, L. Schmidt-Thieme, & J. Kunze (Eds.), *Challenges at the interface of data analysis, computer science, and optimization* (pp. 495–501). Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-64224466-7_50.

Henseler, J., & Schuberth, F. (2020). Using confirmatory composite analysis to assess emergent variables in business research. *Journal of Business Research, 120*, 147–156. https://doi.org/10.1016/j.jbusres.2020.07.026

Henseler, J., R. Sinkovics, R.-J. B. J., Daekwan Kim, R., Ringle, C. M., & Sarstedt, M. (2016). Testing measurement invariance of composites using partial least squares. *International Marketing Review, 33*(3), 405-431. https://10.1108/imr-09-2014-0304.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70. http://www.jstor.org/stable/4615733.

Huaman-Ramirez, R., & Merunka, D. (2019). Brand experience effects on brand attachment: The role of brand trust, age, and income. *European Business Review, 31*(5), 610–645. https://doi.org/10.1108/EBR-02-2017-0039

Keil, M., Tan, B. C. Y., Wei, K.-K., Saarinen, T., Tuunainen, V., & Wassenaar, A. (2000). A cross-cultural study on escalation of commitment behavior in software projects. *MIS Quarterly, 24*(2), 299–325. https://doi.org/10.2307/3250940

Klesel, M., Schuberth, F., Henseler, J., & Niehaves, B. (2019). A test for multigroup comparison using partial least squares path modeling. *Internet Research, 29*(3), 464–477. https://doi.org/10.1108/IntR-11-2017-0418

Klesel, M., Schuberth, F., Niehaves, B., & Henseler, J. (2022). Multigroup analysis in information systems research using PLS-PM: A systematic investigation of approaches. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 53*(3), 26–48. https://doi.org/10.1145/3551783.3551787

Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal, 28*(1), 227–261. https://doi.org/10.1111/isj.12131

Lee, R., & Lockshin, L. (2021). Halo Effects of Tourists' Destination Image on Domestic Product Perceptions. *Australasian Marketing Journal, 19*(1), 7–13. https://doi.org/10.1016/j.ausmj.2010.11.004

Liengaard, B. D., Sharma, P. N., Hult, G. T. M., Jensen, M. B., Sarstedt, M., Hair, J. F., & Ringle, C. M. (2021). Prediction: Coveted, yet forsaken? Introducing a cross-validated predictive ability test in partial least squares path modeling. *Decision Sciences, 52*(2), 362–392. https://doi.org/10.1111/deci.12329

Lim, X. J., Cheah, J. H., Ng, S. I., Basha, N. K., & Liu, Y. (2021). Are men from Mars, women from Venus? Examining gender differences towards continuous use intention of branded apps. *Journal of Retailing and Consumer Services, 60*, Article 102422. https://doi.org/10.1016/j.jretconser.2020.102422

MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology, 90*(4), 710–730. https://doi.org/10.1037/0021-9010.90.4.710

Mahmoud, A. B., Grigoriou, N., Fuxman, L., Reisel, W. D., Hack-Polay, D., & Mohr, I. (2020). A generational study of employees' customer orientation: A motivational viewpoint in pandemic time. *Journal of Strategic Marketing, 30*(8), 746–763. https://doi.org/10.1080/0965254X.2020.1844785

Matthews, L. (2017). Applying Multigroup Analysis in PLS-SEM: A Step-by-Step Process. In H. Latan, & R. Noonan (Eds.), *Partial Least Squares Path Modeling* (pp. 219–243). Springer.

Matthews, L., Hair, J. F., & Matthews, R. (2018). Pls-Sem: The Holy Grail For Advanced Analysis. *Marketing Management Journal, 28*(1), 1–13. https://doi.org/10.1108/IMR-09-2014-0304

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2020). The role of information governance in big data analytics driven innovation. *Information & Management, 57*(7), Article 103361. https://doi.org/10.1016/j.im.2020.103361

Moore, Z., Harrison, D. E., & Hair, J. (2021). Data quality assurance begins before data collection and never ends: What Marketing researchers absolutely need to remember. *International Journal of Market Research, 63*(6), 693–714. https://doi.org/10.1177/14707853211052183

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* ((3rd ed.).). McGraw-Hill.

Pitman, E. J. G. (1938). Significance tests which may be applied to samples from any population. *Journal of the Royal Statistical Society Supplement, 44*(1), 119–130.

Rademaker, M. E., & Schuberth, F. (2021). cSEM: Composite-based Structural Equation Modeling (Version: 0.4.0.9000). https://m-e-rademaker.github.io/cSEM/.

Ramli, N. A., Latan, H., & Solovida, G. T. (2019). Determinants of capital structure and firm financial performance—A PLS-SEM approach: Evidence from Malaysia and Indonesia. *The Quarterly Review of Economics and Finance, 71*, 148–160. https://doi.org/10.1016/j.qref.2018.07.001

Ringle, C., Wende, S., & Becker, J.-M. (2015). SmartPLS3. Bönningstedt: SmartPLS. Retrieved from http://www.smartpls.com.

Roberts, N., & Thatcher, J. (2009). Conceptualizing and testing formative constructs: Tutorial and Annotated Example. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 40*(3), 9–39. https://doi.org/10.1145/1592401.1592405

Roy, S. K., Balaji, M. S., Soutar, G., Lassar, W. M., & Roy, R. (2018). Customer engagement behavior in individualistic and collectivistic markets. *Journal of Business Research, 86*, 281–290. https://doi.org/10.1016/j.jbusres.2017.06.001

Sanchez-Franco, M. J., & Roldán, J. L. (2010). Expressive aesthetics to ease perceived community support: Exploring personal innovativeness and routinised behaviour as moderators in Tuenti. *Computers in Human Behavior, 26*(6), 1445–1457. https://doi.org/10.1016/j.chb.2010.04.023

Sarstedt, M., & Cheah, J.-H. (2019). Partial least squares structural equation modeling using SmartPLS: A software review. *Journal of Marketing Analytics, 7*(3), 196–202. https://doi.org/10.1057/s41270-019-00058-3

Sarstedt, M., Radomir, L., Moisescu, O. I., & Ringle, C. M. (2022c). Latent class analysis in PLS-SEM: A review and recommendations for future applications. *Journal of Business Research, 138*, 398–407. https://doi.org/10.1016/j.jbusres.2021.08.051

Sarstedt, M., & Ringle, C. M. (2010). Treating unobserved heterogeneity in PLS path modeling: A comparison of FIMIX-PLS with different data analysis strategies. *Journal of Applied Statistics, 37*(8), 1299–1318. https://doi.org/10.1080/02664760903030213

Sarstedt, M., Hair, J. F., Pick, M., Liengaard, B. D., Radomir, L., & Ringle, C. M. (2022a). Progress in partial least squares structural equation modeling use in marketing research in the last decade. *Psychology & Marketing, 39*(5), 1035–1064. https://doi.org/10.1002/mar.21640

Sarstedt, M., Hair, J. F., Jr., & Ringle, C. M. (2022b). "PLS-SEM: Indeed a silver bullet"–retrospective observations and recent advances. *Journal of Marketing Theory and Practice*.

Sarstedt, M., Henseler, J., & Ringle, C. M. (2011). Multigroup Analysis in Partial Least Squares (PLS) Path Modeling. *Alternative Methods and Empirical Results., 22*, 195–218. https://doi.org/10.1108/s1474-7979(2011)0000022012

Schade, M., Hegner, S., Horstmann, F., & Brinkmann, N. (2016). The impact of attitude functions on luxury brand consumption: An age-based group comparison. *Journal of business research, 69*(1), 314–322. https://doi.org/10.1016/j.jbusres.2015.08.003

Schlägel, C., & Sarstedt, M. (2016). Assessing the measurement invariance of the four-dimensional cultural intelligence scale across countries: A composite model approach. *European Management Journal, 34*(6), 633–649. https://doi.org/10.1016/j.emj.2016.06.002

Sharma, P. N., Liengaard, B. D. D., Hair, J. F., Sarstedt, M., & Ringle, C. M. (2022). Predictive model assessment and selection in composite-based modeling using PLS-SEM: Extensions and guidelines for using CVPAT. *European Journal of Marketing*.

Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J.-H., Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. *European Journal of Marketing, 53*(11), 2322–2347. https://doi.org/10.1108/ejm-02-2019-0189

Streukens, S., & Leroi-Werelds, S. (2016). Bootstrapping and PLS-SEM: A step-by-step guide to get more out of your bootstrap results. *European Management Journal, 34*(6), 618–632. https://doi.org/10.1016/j.emj.2016.06.003

Ting, H., Fam, K. S., Hwa, J. C. J., Richard, J. E., & Xing, N. (2019). Ethnic food consumption intention at the touring destination: The national and regional perspectives using multigroup analysis. *Tourism Management, 71*, 518–529. https://doi.org/10.1016/j.tourman.2018.11.001

Trojanowski, M., & Kułak, J. (2017). The Impact of Moderators and Trust on Consumer's Intention to Use a Mobile Phone for Purchases. *Journal of Management and Business Administration. Central. Europe, 25*(2), 91–116. https://doi.org/10.7206/jmba.ce.2450-7814.197

Urbach, N., & Ahlemann, F. (2010). Structural equation modeling in information systems research using partial least squares. *Journal of Information Technology Theory and Application (JITTA), 11*(2), 2. https://aisel.aisnet.org/jitta/vol11/iss2/2.

Velayutham, S., Aldridge, J. M., & Fraser, B. (2012). Gender Differences in Student Motivation And Self-Regulation In Science Learning: A Multi-Group Structural Equation Modeling Analysis. *International Journal of Science and Mathematics Education, 10*(6), 1347–1368. https://doi.org/10.1007/s10763-012-9339-y

Voorhees, C. M., Brady, M. K., Calantone, R., & Ramirez, E. (2015). Discriminant validity testing in marketing: An analysis, causes for concern, and proposed remedies. *Journal of the Academy of Marketing Science, 44*(1), 119–134. https://doi.org/10.1007/s11747-015-0455-4

Yáñez-Araque, B., Sánchez-Infante Hernández, J. P., Gutiérrez-Broncano, S., & Jiménez-Estévez, P. (2021). Corporate social responsibility in micro-, small- and medium-sized enterprises: Multigroup analysis of family vs. nonfamily firms. *Journal of Business Research, 124*, 581–592. https://doi.org/10.1016/j.jbusres.2020.10.023

Jun-Hwa Cheah (Jacky) is an Associate Professor at Norwich Business School, the University of East Anglia. His areas of interest include consumer behaviour, quantitative research, and methodological issue. His publications appear in journals such as the European Journal of Marketing, Journal of Retailing and Consumer Services, Marketing Intelligence and Planning, Asia Pacific Journal Marketing and Logistics, Technological Forecasting and Social Change, Total Quality Management and Business Excellent, Management Decision, Internet Research, Information Systems Management, Industrial Management and Data Systems, Tourism Management, Tourism Economics, International Journal of Contemporary Hospitality Management, and etc. He has also received several research awards.

Suzanne Amaro has a PhD in Marketing and Strategy and is an associate professor at the Higher School of Technology and Management of the Polytechnic Institute of Viseu in Portugal. She is currently Vice-President of this higher education school and head of the MSc Marketing Degree. Her current research interests include consumer behaviour, partial least squares and e-Tourism. She has published articles on these topics in top-tier journals. She is a full member of CISED - Centre for Research in Digital Services research unit and a member of CITUR – Centre for Tourism Research, Development and Innovation.

José L. Roldán is a Professor of Management in the Department of Business Administration and Marketing, Universidad de Sevilla, Spain. He holds a Ph.D. His research interests include technology acceptance models, knowledge management, organizational culture, organizational agility, and partial least squares (PLS). His recent contributions have been published in Decision Sciences, Internet Research, Journal of Travel Research, European Journal of Operational Research, International Journal of Project Management, British Journal of Management, Journal of Business Research, European Journal of Information Systems, Computers in Human Behavior, and Industrial Marketing Management, among others. He is currently on the editorial board of the Data Base for Advances in Information Systems. He has also served as Guest Editor of the European Journal of Information Systems, Journal of Business Research, European Management Journal, and International Journal of Sports Marketing and Sponsorship. In addition, Dr. Roldan has been conference co-chair of the 2nd International Symposium on Partial Least Squares Path Modeling—The Conference for PLS Users (June 16–19, Seville, Spain) and the 9th International Conference on PLS and Related Methods (PLS'17) (June 17–19, 2017, Macau, China).