RIGOROUS TESTING OF CAUSALITY MODELS IN CONSUMER BEHAVIOR RESEARCH

PhD Thesis submitted to the Faculty of Economics and Business

Enterprise institute

University of Neuchâtel

For the degree of PhD in Management

by

Eric RAKOTOASIMBOLA

Accepted by the dissertation committee:

Sam Blili, Professor, University of Neuchâtel, Thesis Director

Valéry Bezençon, Professor, University of Neuchâtel, Jury President

Mehdi Farsi, Professor, University of Neuchâtel, Internal Examiner

Dorota Leszczynska, Professor, Ecole de commerce IDRAC, GREDEG CNRS Sophia Antipolis, France, External Examiner

Miguel Palacios, Professor, ESCP Europe Business School, Madrid, Spain**,** External Examiner

Defended on 5th September 2017

Faculté des Sciences Economiques
Avenue du 1er-Mars 26
CH-2000 Neuchâtel
www.unine.ch/seco

LETTRES ET SCIENCES HUMAINES
SCIENCES
DROIT
SCIENCES ECONOMIQUES
THEOLOGIE

UNIVERSITÉ DE NEUCHÂTEL

IMPRIMATUR POUR LA THÈSE

Rigorous Testing of Causality Models in Consumer Behavior Research

Eric RAKOTOASIMBOLA

UNIVERSITÉ DE NEUCHÂTEL

FACULTÉ DES SCIENCES ÉCONOMIQUES

La Faculté des sciences économiques,
sur le rapport des membres du jury

Prof. Sam Blili (directeur de thèse, Université de Neuchâtel)
Prof. Valéry Bezençon (président du jury, Université de Neuchâtel)
Prof. Mehdi Farsi (Université de Neuchâtel)
Prof. Dorota Leszczynska (Ecole de commerce IDRAC – GREDEG CNRS Sophia Antipolis, France)
Prof. Miguel Palacios (ESPC Europe Business School, Madrid)

Autorise l'impression de la présente thèse.

Neuchâtel, le 30 octobre 2017

Le doyen

Mehdi Farsi

Praise be to God!

ACKNOWLEDGMENTS

This dissertation would not have been possible without the help of numerous individuals, who I would now like to thank.

I owe my deepest gratitude to my thesis director, Prof. Sam Blili, for his sound advice and constructive input, for making himself available to me, and for the endless patience that he has shown me. Prof. Sam Blili has taught me more than I could ever have imagined, not only for research, but also about life. He is a great man.

I also express my sincere thanks to Prof. Valéry Bezençon for having agreed to preside over my thesis defense as well as for taking the time to organize it. It is an honor for me to have you as the jury president.

I would like to show my gratitude to the members of my dissertation committee, Prof. Dorota Leszczynska, Prof. Miguel Palacios and Prof. Mehdi Farsi, for having agreed to evaluate my thesis. Thank you for your comments and suggestions, which have greatly improved my work. I feel honoured to have had you sit on my committee.

I also would like to mention and thank the following researchers at the University of Neuchâtel who graciously allowed me to complete my survey by granting me access to their students: Dr. Nolwenn Bühler, Dr. Laurent Moccozet, Prof. Daniel Kraus, Prof. Sam Blili, Dr. Praz Christophe, Prof. Blaise Carron, Prof. Benoit Grevisse, Prof. Anne Prévot, Dr. Christina Grisot, Prof. Kuhn André and Prof. Olivier Crevoisier. In addition, my thanks go out to the students themselves for taking the time to fill out my survey. I am grateful as well to Dr. Suzana Zink and Dr. Katarzyna Jagodzinska for their translation expertise, which helped me improve the language of my survey.

My thanks are also due to the member of the Entreprise Institute for providing me with a warm working environment, and especially to Philippe and Oussama, two extraordinary colleagues. Of course, I am especially grateful for the generous administrative support that I received from Sara over the past years.

I would like to thank the University of Neuchâtel for providing me with the educational infrastructure that has allowed me to write this dissertation.

My thanks go out as well to my parents, Adolphe and Héloïse and my siblings, Eddy, Fabrice, and Fanny, for their unconditional support over these past years. Also thanks to my Uncle Patrick and Aunt Emérentienne who made it possible for me to pursue my studies in Switzerland.

My acknowledgments would not be complete, however, if I did not thank my beautiful wife, Linda, for her support and her patience over the past several years. A busy mother with her own pursuits, she never got tired of listening to me talk about my dissertation. For this, I am eternally grateful.

SUMMARY

The hypothetico-deductive method is one of the pillars of scientific research. Within consumer behavior research in marketing, Structural Equation Modeling (SEM) is an increasingly popular statistical approach to deduction, especially given recent software packages that simplify its use. However, facile reliance on the convenience of SEM can create significant pitfalls, particularly regarding its implementation as a statistical tool. To address this issue, this thesis presents three essays on applying SEM more consistently and rigorously. The first essay proposes a decision protocol for using SEM. The second essay assesses the impact of sample size, choice of estimation methods, and degree of nonnormality of variables on fit indices. Finally, the third essay compares the results of two SEM methods, namely confirmatory factorial analysis (CFA) and exploratory structural equation modeling (ESEM), through four involvement measurement models.

The general purpose of this study is to assist researchers — especially novices in statistics — in applying SEM, while discouraging those methods that have been shown unsuitable in the literature. The decision protocol suggested in the first essay can also be used by journal reviewers as a benchmark for evaluating SEM-based articles. Specifically, for this essay we reviewed four articles on SEM review practices, and formulated a proposed baseline based on a synthesis of the authors' experiences and recommendations.

The second essay emphasizes the importance of choosing the proper fit index when evaluating a model. Mittal and Lee's (1989) causality model of consumer involvement was taken as a model of study. This study demonstrated that the data characteristics (sample size and degree of nonnormality of the variables) and choice of estimation method (maximum likelihood or general least squares methods) affect the fit indices. However, the impacts of each fit index are different. Using the Monte Carlo simulation method, we propose several suggested criteria to assess the Mittal and Lee's model (1989) for other replication studies. Overall, we recommend the use of Standardized Root Mean Square Residual (SRMR) and Root Mean Square Error of Approximation (RMSEA), but also we suggest Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) under certain specific conditions, depending on the sample size and the choice of estimation methods. Adjusted Goodness-of-FIt statistic (AGFI) and Goodness-of-FIt statistic (GFI) are not recommended.

The third essay compares two SEM methods, namely Confirmatory Factor Analysis (CFA) and Exploratory Structural Equation Modeling (ESEM). ESEM is a new approach within SEM, and to our knowledge it has not yet been applied in the field of consumer behavior research in Marketing. We compare these two methods by using the concept of consumer involvement in terms of footwear products. This concept was chosen because the issue of its operationalization has caused much controversy in the literature. After comparing four main models used to measure consumer involvement in CFA and ESEM, we rank their performance according to their psychometric qualities. To perform this analysis, data were gathered by means of a survey among the students of the University of Neuchâtel. The results of the study show that the two methods are not interchangeable, but that ESEM can be a helpful complementary tool to CFA for operationalizing constructs.

Keywords: *structural equation modeling, psychometric qualities, fit indices, Monte Carlo simulation, confirmatory factor analysis, exploratory structural equation modeling, measurement models, consumer involvement.*

RÉSUMÉ

Les chercheurs de la discipline de l'étude du comportement du consommateur en Marketing utilisent couramment la méthode hypothético-déductive pour fournir une explication causale des phénomènes. Les Méthodes des Équations structurelles (MES) sont parmi les méthodes qui adhèrent à cette pratique et elles constituent des outils privilégiés dans la discipline de l'étude du comportement du consommateur en Marketing. La vulgarisation des logiciels facilitant le recours aux MES ont permis d'accélérer leur utilisation dans le domaine de cette étude. Toutefois, cette facile utilisation des MES ne s'est pas faite sans revers, notamment sur sa mise en œuvre en tant qu'outil statistique. Au vu de ces constats, cette thèse présente trois essais sur des recommandations d'application rigoureuse de la MES. Le premier essai est consacré à une proposition de protocole de décision pour l'utilisation de la MES. Le deuxième essai présente l'impact de la taille de l'échantillon, du degré de nonnormalité des variables et du choix des méthodes d'estimations sur les indices d'ajustement pour tester un modèle. Le troisième essai compare les résultats de deux méthodes d'équations structurelles, l'analyse factorielle confirmatoire (AFC) et la méthode exploratoire par les équations structurelles (MEES), à partir de quatre modèles de mesures de l'implication du consommateur.

Le but de cette étude est de proposer aux chercheurs, et plus particulièrement aux chercheurs non chevronnés des statistiques, des solutions pour surmonter les mauvaises pratiques identifiées dans la littérature pour l'application de la MES.

À ce titre, le protocole de décision présenté dans le premier essai peut être également utilisé par les réviseurs d'articles scientifiques en tant qu'outil d'étalonnage pour évaluer les pratiques d'application de la MES. Pour ce premier essai, nous nous sommes basés sur quatre articles d'évaluations des pratiques de la MES. Ces quatre articles, nous ont permis à priori de formuler une base de synthèse des recommandations et des expériences pour construire le protocole de décision.

Le deuxième essai présente l'importance du choix des indices d'ajustements pour évaluer la qualité d'un modèle. Le modèle de causalité de Mittal et Lee (1989) sur l'implication du consommateur a été pris comme modèle d'étude. Par le truchement de la méthode de simulation de Monte Carlo,

l'étude a révélé que les caractéristiques ayant trait aux données (la taille de l'échantillon et le degré de non-normalité des variables) et au choix de la méthode d'estimation affectent les indices et les interprétations qui en découlent. Les résultats de cette étude suggèrent l'utilisation des indices tels que le SRMR (Standardized Root Mean Square Residual) et le RMSEA (Root Mean Square Error of Approximation) pour évaluer de futures réplications du modèle de Mittal et Lee (1989). Les indices CFI (Comparative Fit Index) et TLI (Tucker-Lewis Index) sont également recommandés à condition que certaines conditions spécifiques à la taille de l'échantillon et au choix de la méthode d'estimation soient respectées. Finalement, les recours aux indices AGFI (Adjusted Goodness-of-FIt) et GFI (Goodness-of-FIt) ne sont pas conseillés.

En ce qui concerne le troisième essai, la MEES est une nouvelle approche des MES. À notre connaissance, aucune application de cette nouvelle approche n'a été faite au sein de la discipline du comportement du consommateur en Marketing. La comparaison des deux méthodes est faite au moyen du concept de « l'implication du consommateur » pour le produit chaussure. Le choix s'est porté sur ce concept du fait qu'il suscite beaucoup de controverse dans la littérature concernant son opérationnalisation. En comparant quatre principaux modèles de mesures de l'implication du consommateur, nous avons fourni une classification de leur performance au travers des qualités psychométriques. Pour effectuer cette analyse, une collecte de données par enquête a été effectuée auprès des étudiants de l'université de Neuchâtel. Les résultats de l'étude ont montré que les deux méthodes ne sont pas substituables mais sont plutôt complémentaires.

Keywords: *modélisation par équations structurelles, qualités psychométriques, indices d'ajustement, simulation de Monte Carlo, analyse factorielle confirmatoire, méthodes exploratoires par les équations structurelles, modèles de mesure, implication du consommateur.*

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ADF	Asymptotic distribution-free
AGFI	Adjusted goodness-of-fit statistic
AIC	Akaike information criterion
A-I-C	Antecedents-implication-consequences
ANOVA	Analysis of variance
BIC	Bayesian information criterion
CFA	Confirmatory factorial analysis
CFI	Comparative fit index
CIP	Consumer involvement profiles
CVI	Cross validation index
EFA	Exploratory factorial analysis
EICR	Extended individual case residuals
EPC	Expected change statistics
ESEM	Exploratory structural equation modeling
FIML	Full information maximum likelihood
GFI	Goodness-of-fit statistic
GLS	Generalized least squares
ICOMP	Information complexity criterion
LISREL	Linear structural relations
LR	Likelihood ratio
MAR	Missing at random
MCAR	Missing completely at random
MI	Modification indices
ML	Maximum likelihood
MLM	Maximum likelihood parameter estimates with standard errors and a mean adjusted Chi-square test statistic
MLMV	Maximum likelihood parameter estimates with standard errors and a mean- and variance-adjusted Chi-square test statistic
MLR	Maximum likelihood estimation with robust standard errors

MTMM	Multitrait multimethod matrix
NMAR	Not missing at random
PII	Product involvement inventory
PIIA	Product involvement inventory for advertisement
RMSEA	Root mean square error of approximation
RMR	Root mean square residual
RRPII	Revised - revised product involvement inventory
SAS	Statistical Analysis System
SEM	Structural equation modeling
SPSS	Statistical Package for the Social Sciences
SRMR	Standardized root mean square residual
TLI	Tucker-Lewis index
WLSM	Weighted least squares mean-adjusted
WLSMV	Weighted least squares mean-adjusted and variance-adjusted

CHAPTER I. INTRODUCTION

1.1 CONTEXT OF THE RESEARCH

We know that one of the main expectations in consumer behavior research in marketing is being able to explain and predict consumer behavior (Howard & Sheth, 1969; Solomon, Bamossy, Askegaard & Hogg, 2016). In this study, we focus on the application of covariance based structural equation modeling (CB-SEM). We follow the normal practice of calling this model simply SEM, though partial-least square SEM (PLS-SEM) is technically also a type of SEM. The difference is that CB-SEM is more about explaining the covariation among the observed and latent variables. It aims at minimizing bias and providing tests of model fit (Evermann & Tate, 2016). By contrast, PLS-SEM is more about prediction-oriented goals and aims at maximizing the explained variance of all dependent variables (Evermann & Tate, 2016).

The precise origins of the SEM method are unclear. According to Aish-Van Vaerenbergh (1997), analysis of covariance structure arose from the convergence of work in several independent fields: psychometrics with factor analysis, biometrics, sociology with path analysis, and econometrics with systems of simultaneous equations. In the 1970s, the development of the software LISREL (Linear Structural RELations) by Karl Jöreskog popularized the method within psychology, while Richard Bagozzi introduced it within marketing research (Baumgartner & Homburg, 1996). Since then, many studies of consumer behavior have utilized SEM, but with significant annual variations. (e.g., Richter, Sinkovics, Ringle, & Schlägel, 2016). It is known that SEM provides some advantages over classical statistical methods such as multiple regression analysis, canonical correlation analysis, and variance analysis (e.g., Bagozzi & Yi, 2012). For example, SEM enables simultaneous testing for the existence of various causal links among several independent and dependent variables, single-tool verification of the psychometric qualities of measurement models, and examination of models while considering measurement errors. However, as SEM is mostly used with cross-sectional data it needs more conditions to address causality, compared to experimental data.

Furthermore, like any other statistical modeling technique, SEM requires applying some conditions such as a reasonable sample size or distributional assumptions on the data. Although these conditions are of vital importance they are not always verified, so the statistical methods adopted are often ill-suited to their tasks (e.g., Martínez-López, Gázquez-Abad, & Sousa, 2013) . They must be properly applied to prevent the research from yielding an erroneous conclusion (Richter, Sinkovics, Ringle, & Schlägel, 2016). For example, hypotheses might be rejected because the sample size was not sufficient to apply the selected method. Such errors would not only delegitimize that study, but also weaken any later research that extended or otherwise relied on the flawed study. Along similar lines, poorly executed SEM research could lead practitioners to make incorrect decisions, which might be costly for their organizations.

Within the field of marketing, there are only a few studies on structural equation modeling review practice. We only identified four articles from 1970 to 2016: Baumgartner and Homburg (1996); Hulland, Chow, and Lam (1996); Martínez-López et al. (2013) and Richter et al. (2016). The results of these studies show that every application of SEM can be potentially impacted by different problems. Some of these problems are old and well-known in the literature, such as the problem of degree of nonnormality. In fact, in social science data frequently violate the hypothesis of normality (Harlow, 1985), and only 6.5% of published studies specifically report treating this problem (Martínez-López et al., 2013). This situation casts doubt on the methods used to address nonnormality, such as the estimation method. Moreover, it has been recognized in the literature that the standard likelihood estimation methods are impacted by degree of nonnormality (e.g., Olsson, Foss, Troye, & Howell, 2000).

Other problems, such as the misspecification of measurement models, were also identified. Although recommendations for the choice of reflective and formative measurement models were proposed in the literature more than 10 years ago, it seems that the problem is still on the agenda (see Jarvis, MacKenzie, & Podsakoff, 2003). For example, 20% of research articles do not provide a justification for the type of measurement model (Richter et al., 2016). These cases are simply examples; other problems do exist (see Hulland et al., 1996; Martínez-López et al., 2013; Richter et al., 2016).

According to Steenkamp and Baumgartner (2000), the use of SEM seems to be peripheral in marketing research. This suggests that few marketing scholars try to understand the method more deeply. Moreover, most methodological articles on SEM are not published in marketing journals, and can indeed be difficult to understand without a quantitative background (Steenkamp & Baumgartner, 2000). For example, the prestigious journal *Structural Equation Modeling: a Multidisciplinary Journal* is more oriented to mathematics disciplines than most other interdisciplinary journals. This situation compounds the complexity of understanding the rigorous application of SEM, especially for novice researchers.

Furthermore, marketing journal editors rarely insist that authors address issues with reporting SEM applications, which may lead to more bad practices (Babin, Hair, & Boles, 2008). Indeed, many journals require researchers to follow a publication template, which leaves little or no space for reporting on SEM practice. Although we do not question these editors' expertise or the quality of their journals, to ensure intellectual progress it is important to give the necessary data to replicate published SEM results. This situation complicates the rigorous evaluation of SEM-based articles. It also prevents improvements in the use of SEM, because techniques in the domain evolve and the results can differ from one technique to another.

1.2 PURPOSE AND RESEARCH STRUCTURE

This study is composed of three independent essays, each with its own research design. All three essays are conceptually independent, but share the common purpose of improving the use of SEM in the scientific community.

The first essay is entitled « Towards a decision protocol for research models based on structural equation modeling ». This study focuses on a synthesis of recommendations and experiences of SEM usage in marketing journals. SEM is used in many social science disciplines. As it is difficult to form a transdisciplinary synthesis of SEM usage, we limited our research to articles in marketing journals that reviewed SEM practices. We used Harzing's Publish or Perish software to search for articles. For the keywords: *recommendations, SEM, best practices, fit indices, measurement model, estimation method,* and *review* the software generated a list of 1000 articles published from 1970 to 2015. Among this list, only four articles reviewed all the different steps to apply SEM. The other articles examined specific steps in the process. These four articles were used as a starting point to build our synthesis of recommendations and experiences.

The second essay is entitled « Impact of sample size, nonnormality of variables and estimation method on fit indices: application to the causal model of consumer involvement ». This study evaluates the impact of sample size, degree of nonnormality, and choice of estimation methods on fit indices. We chose Mittal and Lee (1989) causal involvement model for the simulation model.

This model is also known as Antecedents-Involvement-Consequences (A-I-C) model. Ideally, we wished to obtain a model which was tested by SEM to replicate the results and to apply alternative techniques that might be more appropriate. However, we did not receive any positive answers from our 150 e-mails. Various explanations were given such as confidentiality, ongoing projects, and loss of data. Thus, we decided to do a Monte Carlo simulation to assess the impact of the three variables on fit indices. There are three main reasons for the choice of Mittal and Lee (1989) causal involvement model. First, Mittal and Lee (1989) did not confirm the model with their data. A replication of the model by Flynn and Goldsmith (1993) gave the same conclusion. However, some replications from other researchers produced contradictory results (e.g., Bezençon & Blili, 2010). So, we decided to investigate whether the difference lies in the use of SEM rather than in the data. Second, as the causal involvement model is important to consumer behavior research in Marketing (e.g., Solomon et al., 2016) we deemed it appropriate to use this model for our simulation. Third, Mittal and Lee (1989) provided well-estimated parameters of the test of the model. This allowed us to consider these parameters as population values for our simulation.

The third essay is entitled « Confirmatory factor analysis and exploratory structural equation modeling: the case of an assessment of four measurement models of consumer involvement ». This study compares two methods of structural equation modeling, namely confirmatory factorial analysis (CFA) and exploratory structural equation modeling (ESEM). ESEM is well- known in psychological and educational studies (e.g., Marsh, Liem, Martin, Morin, & Nagengast, 2011). As psychology and marketing are closely related disciplines, it is also relevant to apply ESEM to marketing research to assess measurement models. Empirical application of the comparison consists of comparing four models for measuring consumer involvement. These comparisons are based on several psychometric criteria: reliability, convergent validity, discriminant validity, and nomological validity. Consumer involvement measurement models were chosen because of the importance of involvement in consumer behavior research in Marketing (e.g., Solomon et al., 2016), and because these models have been subject to controversies regarding their dimensional structure (e.g., Zaichkowsky, 1994; McQuarrie & Munson, 1992). The four measurement models were selected from the *Handbook of Marketing Scales* (Bearden, Netemeyer & Haws, 2011)

Two of the four models, namely consumer involvement profile CIP (Laurent & Kapferer, 1985) and personal involvement inventory PII (Zaichkowsky, 1985), are the most-cited models in the literature. Both of them are also foundational to later models such as the personal involvement inventory to advertising PIIA (Zaichkowsky, 1994), the revised product involvement inventory RRPII (McQuarrie & Munson, 1992), and modified PII (Mittal, 1995). The article of Zaichkowsky (1985) on original PII is classified as the 4th most-cited article in the highly-ranked *Journal of Consumer Research* (JCR) from 1974 to 2014 (Wang, Bendle, Mai, & Cotte, 2015). JCR is ranked in the first quartile (19 of 121) by Thomson Reuters. The article has 6742 citations according to Publish or Perish software. CIP (Laurent & Kapferer, 1985) is also a relatively well-cited model for measuring consumer involvement in products (3003 citations). It was published in the highly-ranked Journal of Advertising (25/121, according to Thomson Reuters). The PIIA (Zaichkowsky, 1994), RRPPII (McQuarrie & Munson, 1992), and modified PII (Mittal, 1995) measurement models are included in our study because they are the revised versions of PII. In order to compare the CFA and ESEM methods, data were gathered by means of a survey among the students of the University of Neuchâtel.

1.3 RESEARCH QUESTIONS

This study addresses the following research questions:

- **ESSAY I:** How can we improve SEM recommendations and practices in consumer behavior research in Marketing?
- **ESSAY II:** How do conclusions based on fit indices vary relative to sample size, estimation method, and degree of nonnormality within the A-I-C model of Mittal and Lee (1989)?
- **ESSAY III**: For assessing the psychometric quality of consumer involvement measurement models, are there different results between Confirmatory Factor Analysis (CFA) and Exploratory Structural Equation Modeling (ESEM)?

1.4 CONTRIBUTIONS OF THE RESEARCH

The first essay contributes to the advancement of marketing research by providing a decision protocol for applying SEM. The decision protocol is simple and intuitive to use, and the flowchart provides a more coherent process. Unlike the tailored case-by-case approach of the previous articles on SEM protocols, our decision protocol does not require the novice researcher to order the steps, which is time-consuming. Moreover, at each stage we provide both explanations of what to do and of how to apply the method. This helps novice users of SEM identify the standard steps they should follow for testing hypotheses and theories. Moreover, our decision protocol covers the broadest variety of topics compared to the other reviews of SEM. Finally, the article provides a synopsis of the neglected and even forgotten, yet fundamental, criteria for the use of SEM.

The second essay assesses the impact of sample size, degree of nonnormality, and estimation method on fit indices. A Monte Carlo simulation was used to assess this impact. Results show the extent to which fit indices can diverge from each other when used to assess a model, and generate new recommendations on choosing fit indices for assessing A-I-C models. Applying the Monte Carlo simulation method in this study shows how we can assess the suitability of fit indices regarding sample size, choice of estimation, and degree of nonnormality of data.

The third essay provides a comparison of two structural equation approaches, ESEM and the CFA, across four consumer involvement models: CIP, PIIA, RRPII, and modified PII. Results shows that the two methods diverge in confirming discriminant validity for PIIA and CIP. This poses the question of which method is reliable. In addition, this study provides a ranking of these four well-known consumer involvement measures. The ranking will help researchers select measurement models based on the performance of each measure on each dimension of the essay's psychometric criteria: reliability, convergent validity, discriminant validity, and nomological validity.

1.5 EPISTEMOLOGICAL APPROACH

In Popper's (1962) argument against the verifiability of theories, he asserts that it is not possible to test theories in science and to establish their truth. The only thing we can do is to test whether a theory is false by showing the consequences of the theory to be false. As is known, the form of the argument against a hypothesis would be:

> *"If H is true, then C is true.*
>
> *C is false*
>
> ______________
>
> *Therefore, H is false."(Mulaik & James, 1995, p. 121).*

This is a form of the hypothetico-deductive method which can be applied to SEM. SEM models fall under the positivist approach, where phenomena are considered to exist and independently of the observer (Mulaik & James, 1995). We can replicate such models to have the same results under similar conditions. In a sense it is possible to generalize these results, in contrast to the interpretative approach which is context-dependent. So, in this study we adopt a positivist approach for causal explanations.

1.6 REFERENCES

Aish-Van Vaerenbergh, A. (1994). Modèles statistiques et inférences causales : analyse des structures de covariances avec LISREL. In R. Franck, *Faut-il chercher aux causes une raison ? L'explication causale en sciences humaines* (pp. 106-130). Librairie philosophique Vrin. Institut interdisciplinaire d'études épistémologiques.

Babin, B. J., Hair, J. F., & Boles, J. S. (2008). Publishing Research in Marketing Journals Using Structural Equation Modeling. *Journal of marketing theory and practice*, 16(4), 279-286.

Bagozzi, R. P., & Yi, Y. (2012). Specification, evaluation, and interpretation of structural equation models. *Journal of the academy of marketing science*, 40(1), 8-34.

Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International journal of Research in Marketing*, 13(2), 139-161.

Bearden, W., Netemeyer, R. and Haws, K. (2011). *Handbook of marketing scales*. 3rd ed. Thousand Oaks, CA: Sage Publications.

Bezençon, V., & Blili, S. (2010). Ethical products and consumer involvement: what's new? *European Journal of Marketing*, 44(9/10), 1305-1321.

Evermann, J., & Tate, M. (2016). Assessing the predictive performance of structural equation model estimators. *Journal of Business Research*, 69(10), 4565-4582.

Flynn, R. and Goldsmith, R. (1993). A causal model of consumer involvement: Replication and critique. *Journal of Social Behavior and Personality*, 8(6), 129-142.

Harlow, L. (1985). *Behavior of some elliptical theory estimators with nonnormal data in a covariance structures framework: A Monte Carlo study*. Los Angeles: University of California.

Howard, J. and Sheth, J. (1969). *The theory of buyer behavior*. New York: John Wiley and Sons.

Hulland, J., Chow, Y. H., & Lam, S. (1996). Use of causal models in marketing research: A review. *International journal of Research in Marketing*, 13(2), 181-197.

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of consumer research*, 30(2), 199-218.

Laurent, G., & Kapferer, J.-N. (1985). Measuring consumer involvement profiles. *Journal of Marketing Research*, 22, 41-53.

Marsh, H. W., Liem, G. A. D., Martin, A. J., Morin, A. J. S., & Nagengast, B. (2011). Methodological Measurement Fruitfulness of Exploratory Structural Equation Modeling (ESEM): New Approaches to Key Substantive Issues in Motivation and Engagement. *Journal of Psychoeducational Assessment*, 29(4), 322-346.

Martínez-López, F. J., Gázquez-Abad, J. C., & Sousa, C. M. (2013). Structural equation modeling in marketing and business research: Critical issues and practical recommendations. *European Journal of Marketing*, 47(1/2), 115-152.

McQuarrie, E. and Munson, M. (1992). A revised product involvement inventory improve usability and validity. *Advances in Consumer Research*, 19, 108-115.

Mittal, B. (1995). A comparative analysis of four scales of consumer involvement. *Psychology and Marketing*, 12(7), 663-682.

Mittal, B., & Lee, M.-S. (1989). A causal model of consumer involvement. *Journal of Economic Psychology*, 10(3), 363-389.

Mulaik, S. and James, R. (1995). Objectivity and reasoning in science and structural equation modeling. In: R. Hoyle, ed., *Structural equation modeling: Concepts, issues, and applications* (pp.118-137). Thousand Oaks, CA: Sage Publications.

Olsson, U. H., Foss, T., Troye, S. V., & Howell, R. D. (2000). The Performance of ML, GLS, and WLS Estimation in Structural Equation Modeling Under Conditions of Misspecification and Nonnormality. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(4), 557-595.

Popper, K., (1962). *Conjectures and refutations*. The growth of scientific knowledge, New York: Basic Books.

Richter, N., Sinkovics, R., Ringle, C., & Schlägel, C. (2016). A critical look at the use of SEM in international business research. *International Marketing Review*, 33(3), 376-404.

Solomon, M., Bamossy, G., Askegaard, S. and Hogg, M. (2016). *Consumer behaviour*. Harlow, England: Pearson.

Steenkamp, J.-B. E. M., & Baumgartner, H. (2000). On the use of structural equation models for marketing modeling. *International journal of Research in Marketing*, 17(2), 195-202.

Wang, X., Bendle, N. T., Mai, F., & Cotte, J. (2015). The Journal of Consumer Research at 40: A Historical Analysis. *Journal of consumer research*, 42(1), 5-18.

Zaichkowsky, J. L. (1985). Measuring the involvement construct. *Journal of consumer research*, 12(3), 341-352.

Zaichkowsky, J. L. (1994). The Personal Involvement Inventory: Reduction, Revision, and Application to Advertising. *Journal of advertising*, 23(4), 59-70.

CHAPTER II. ESSAY 1: TOWARDS A DECISION PROTOCOL FOR RESEARCH MODELS BASED ON STRUCTURAL EQUATION MODELING

ABSTRACT

This paper offers a synthesis of recommendations and experiences for applying SEM. The procedure is based on a survey of articles published since 1996 containing recommendations for applying structural equations to marketing issues : Baumgartner & Homburg (1996); Chin, Peterson & Brown (2008); Hulland, Chow & Lam, (1996); Martínez-López, Gázquez-Abad & Sousa (2013). The recommendations extracted from these articles were enriched by examining articles resulting from multiple disciplines journals. This paper’s findings have enabled the development of a flowchart featuring a decision protocol that can assist researchers in proceeding with scientific rigor when testing behavioural models based on structural equation modeling. We tried to incorporate recommendations on applying SEM into an integrating framework, with the purpose of guiding researchers through the process of making their decisions, from the beginning to the conclusion of their empirical research. This study focuses on standard SEM methods for cross-sectional data. Accordingly other advanced and very specific SEM methods like longitudinal, experimental, Bayesian and other approaches were not pursued. An eight-step decision protocol is proposed for reference when applying SEM. This system can also be applied in social-science disciplines other than marketing.

Keywords: *structural equation modeling, psychometric qualities, fit indices, formative and reflective models, equivalent models.*

2.1 INTRODUCTION

One of the main expectations and promises of consumer behavior research and practice is understanding and predicting consumer behavior. The importance given to structural equation modeling (SEM) is based on this premise. SEM is about forming better predictions, explanations, and understandings of behavioral phenomena in situations involving adoption or purchase (Chintagunta, Erdem, Rossi, & Wedel, 2006). The use of SEM as a tool for analyzing causal phenomena still raises debates and controversies in the scientific literature (e.g., Bagozzi, 1977; Bollen & Pearl, 2013; Bullock, Harlow, & Mulaik, 1994) but SEM still has a privileged status in the world of research. Indeed, since its implementation in the field of consumer behavior research in Marketing by Bagozzi (1977), the appeal of SEM has grown among researchers publishing in the major marketing journals (Baumgartner & Homburg, 1996; Hulland et al., 1996; Martínez-

López et al., 2013). Technical developments have also emerged to address the methodological weaknesses reported in the literature. However, these recommendations are rarely applied (Baumgartner & Homburg, 1996; Hulland et al., 1996; Martínez-López et al., 2013), which raises doubts about the reliability and validity of SEM-based research.

It is true that it is difficult for a non-statistician scientist to master the extent and complexity of knowledge about SEM that has accumulated over nearly 40 years. Indeed, most recommendations are mentioned only fractionally. There is no holistic framework that considers all the choices forced through the different dimensions of empirical tests. For example, one author may only treat the influence of the sample size, the estimation method, and the degree of complexity of the model on the fit indices (Fan, Thompson, & Wang, 1999), while another may focus on the influence of the number of indicators, the estimation method, and the degree of model complexity (Kenny & McCoach, 2003). In this context, we have tried to incorporate standard recommendations within an integrative framework, guiding the researcher throughout the development of his or her choices, from the beginning to the completion of his or her empirical research.

Some researchers recommend following a sequence of steps for applying SEM (e.g., Chin, Peterson, & Brown, 2008), while others address certain key elements without emphasizing any particular procedure (Baumgartner & Homburg, 1996; Hulland et al., 1996; Martínez-López et al., 2013). Chin et al. (2008) recommend the following six steps for applying SEM: specifying the model, identifying the model, estimating the model, evaluating the model, re-specifying the model, and reporting on the methods used for applying SEM. Hoyle (2012) suggests seven steps: specification, data acquisition and preparation, identification, estimation, evaluation, respecification, and interpretation and reporting. The processes of acquiring, preparing, and identifying data are also incorporated into SEM's application framework, but the author does not specify whether these processes constitute a step in themselves or are simply a consideration that must be borne in mind.

In this paper, following the example of Chin et al. (2008), the identification step is considered a full step, while the estimation step is replaced by an analysis of the data structures. This is because the estimation step is treated as a sub-step, as the estimation method is considered a variable to be chosen depending on the data structure (e.g., sample size, degree of nonnormality). Another step called "generalization of the theoretical model" is incorporated into our integrated framework. In this step, we discuss procedures likely to improve the model's external validity, such as evaluating alternative models or performing cross-validation. Several authors similarly consider such procedures major elements that should be considered when applying SEM (e.g., Martínez-López et al., 2013).

Accordingly, we present a decision protocol comprising eight steps. The steps are itemized as follows:

Step 1: Model specification;
Step 2: Model identification;
Step 3: Analysis of data structure;
Step 4: Empirical validation of psychometric qualities of the measurement variables;
Step 5: Evaluation of the fit of the model;
Step 6: Model respecification;
Step 7: Generalization of the theoretical model; and
Step 8: Reporting.

2.2 RESEARCH METHODS

In the first stage, we delimited our review of articles on SEM methods to marketing journals. To establish a baseline for recommendations, the selected articles concern only SEM review practices. Articles which handle specific parts of applying SEM were not considered. Search keywords for our review includes *recommendations, SEM, best practices, fit indices, measurement model, estimation method*, and *review*. The search was done with Harzing's Publish or Perish software, for the research period from 1970 to 2015. After reviewing the one thousand articles identified by the software, we chose four articles of particular interest: Baumgartner and Homburg (1996), Hulland et al. (1996), Chin et al. (2008) and Martínez-López et al. (2013). Compared to the other articles selected by the software, these articles propose a more complete application of SEM. We decided to use these four articles as the baseline for our synthesis of recommendations and experiences regarding the use of SEM. For each article, we identified all the themes that are addressed regarding SEM application (see Table 1). We assessed each theme based on the level of recommendations. A low level of recommendation was assigned if the article only mentioned what to do. A medium level was given if the article mentioned what to do and gave some synthesized recommendations. A high level was given if the article not only mentioned and synthesized the recommendations but also gave some examples to better illustrate the issues.

In the second stage, we searched for articles in marketing journals which handled some specific parts of SEM that is highly cited in a thematic area, even if that area was also treated in the four baseline articles. For example, we included Diamantopoulos and Winklhofer (2001) article because it was first ranked for the choice of measurement model based on Harzing's Publish or Perish software. In the third stage, we selected books and articles outside marketing that were relevant to our research. Specifically, we searched for documents that treated SEM themes that are rarely treated in marketing research. For example, regarding the theme of equivalent models, we decided to select the work of Hershberger (2006). Our choices for this category were subjective. Finally, in the fourth stage, articles that treated advanced SEM methods were not considered. For example, these articles treated nonlinear structural equation modeling, longitudinal, and Bayesian SEM.

Table 1. Summary of the journal of review practice on SEM in consumer behavior research in Marketing.

AUTHORS	*A*	*B*	*C*	*D*	*E*	*Our synthesis*
Theoretical consideration and model specification		+			+	Study objectives: ➢ Explanatory (development of a theory) and/or ➢ Predictive
				+	+ +	Type of studies: ➢ Confirmatory ➢ Exploratory ➢ Model comparison
			+		+	Theoretical justifications: ➢ Connections of variables postulated by the model ➢ Variables unrelated in the model
					+	Content validity: ➢ Validation by experts ➢ Literature review
			+ + +		+ + +	Ontological status of relations between indicators and constructs: ➢ Reflective ➢ Formative
Model identification	+ +	+ +	+ +	+	+ + +	Suggestions of identification heuristic rules
Analysis of data structures					+ +	Independence of observations
	+	+ +	+		+ + +	Treatment of missing values
	+ + +	+ + +	+ +	+ +	+ + +	Consideration of the nature of distributions of endogenous variables
	+ +	+ +		+ +	+ +	Reliability tests of measurements
	+ + +	+ + +	+ +	+ +	+ + +	Recommendation on the sample size
Empirical validation of the measurement models					+ + +	Distinction between validation of the formative and reflective constructs
	+	+ +		+ +	+ +	Convergent validity and Discriminant validity
		+		+ +	+	Nomological validity

Model estimation	+ +					Inherent problems of non-convergence of optimal solutions due to several factors (e.g. negative variance, sample size, outliers, etc.)
		+ +		+ +	+ + +	Alternative recommendations to address the violation of the normality of data: ➢ Continuous data ➢ Categorical data
Model evaluation	+ + +	+ + +	+ + +	+ + +	+ + +	Recommendations for the selection of indices: ➢ Overall fit indices ➢ Incremental indices ➢ Parsimony indices
	+ +	+ +	+	+ +	+ +	Recommendations on the decision-making threshold value
Model re-specification	+ + +	+ +	+ + +	+ +	+ + +	Recommendations for model re-specification
	+	+	+ +		+ + +	Methods of model re-specification ➢ "*Backward search*" and "*Forward search*" ➢ Other: Statistics-*t*, index modification, EPC ("*Expected Parameter Change*"), etc.
Generalization of the conclusion of research			+	+	+ + +	Evaluation of statistical power
	+ +		+	+ +	+ +	Validation with an independent sample
				+	+ +	Identification of alternative models
Reporting		+ +	+ + +	+ + +	+ +	Recommendations for reporting

+ Low (article only mentioned what to do)

++ Average (article mentioned what to do and gave some synthesized recommendations)

+++ High (article not only mentioned and synthesized the recommendations but also gave some examples)

Author A: Baumgartner and Homburg (1996) - Author B: Hulland et al. (1996) - Author C: Chin et al. (2008) - Author D: Martinèz-Lopèz et al. (2013) - Author E: Authors of the article

2.3 THE DECISION PROTOCOL FOR SEM PROCESS

SEM is a confirmatory approach of a system of hypothetical relationships between variables with the purpose of identifying the parameters of a model capable of reproducing the empirical variance covariance matrix of a sample (Hair, Black, Babin, & Anderson, 2010).

Non-rejection of the model is not a confirmation evidence for validating models. Non-rejection merely means that the model shows a good fit to the data. Stating such a finding is far from an assertion of causality (Biddle & Marlin, 1987; Bullock et al., 1994).

The initial application of SEM requires a causal hypothesis. These hypotheses stem from the previous studies, scientific knowledge, logical arguments, sequential priorities and any additional evidence there may be in support of these factors (Fan et al., 2016). They are also extended to the measurement model, corroborating the link between latent and observed variables.

The step following the causal hypothesis is the model identification. This is a technical test for evaluating whether the equation system is solvable. If the model is unidentified, researchers can in principle use techniques for slightly adjusting the model, provided they entail no risk of changing the initial interpretation.

The following section addresses the steps that require empirical data. Then, we begin analyzing the data. The manner of gathering data is not considered because we only address the process of applying SEM. We will therefore focus more on data analysis, which is a diagnostic step that enables making a selection among the existing statistical testing packages. The analysis focuses on major issues such as sample size, missing values, variable distribution and independence of data. With regard to statistical tests, the aim is to evaluate the model's fit to the data by means of fit indices as well as with structural coefficients. A result indicating poor fit to the data, for example, would call into question the truth of the causal hypothesis formulated. In practice, such a result moves researchers either to evaluate the power of tests or to re-specify their models. The latter method, however, usually entails a switch in method from confirmatory research to exploratory research (Bollen & Pearl, 2013).

Nonetheless, even if the model being studied fits the data well, we cannot therefore conclude that the model is valid. As mentioned earlier, this does not prove causality, but it makes the hypothesis of causation plausible. Such a result must not only be replicated in other samples but must also be compared with other equivalent or alternative models to be generalizable. Until proven otherwise, the results verifying the theory acquire a certain plausibility and may be used to explain the phenomenon being studied (Popper, 1968).

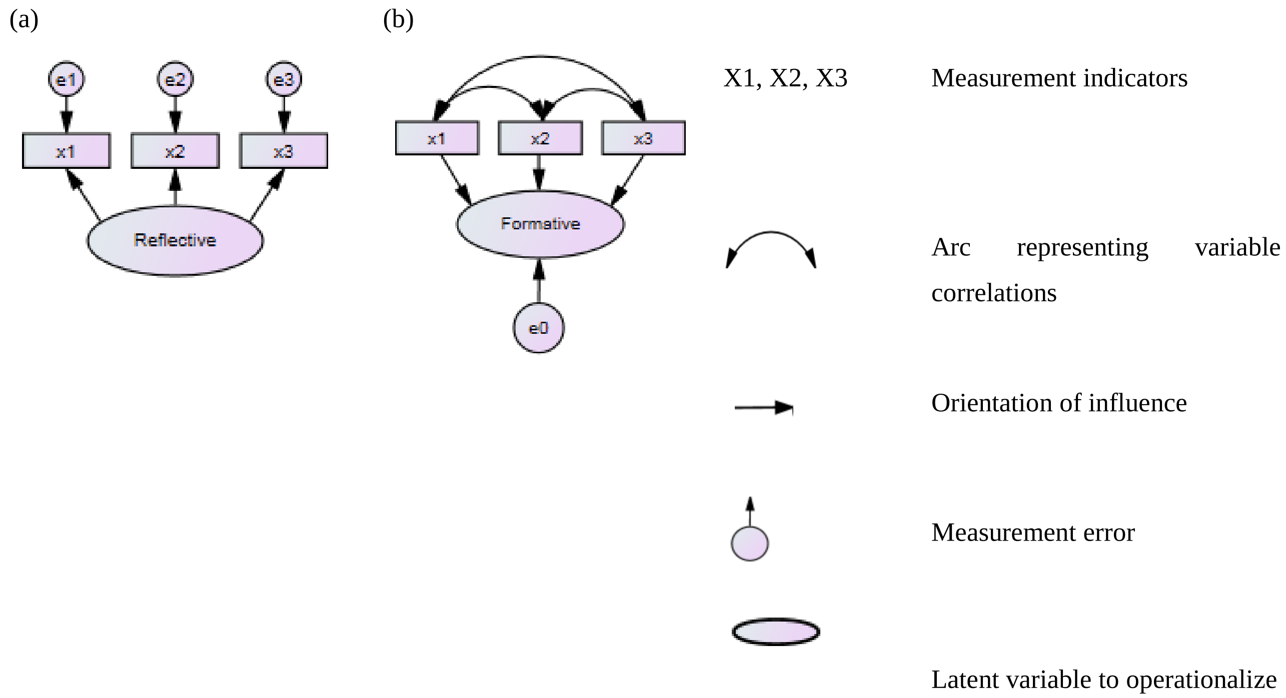
Step 1: Model specification

Specifying a model requires specifying its structural model as well as its measurement models. Structural models contain two or more latent variables. They define the relationship between the latent variables (e.g., unidirectional, bidirectional, correlations). A measurement model shows the link between the latent variable and the groups of indicators or the items that are supposed to measure it.

Step 1.1: Measurement model specification

To evaluate a model's fit to the data, the concepts or constructs used in the model must be observable and measurable. In other words, the concepts or constructs must be operationalized. Operationalizing means designating which measurement indicators are necessary for applying and delimiting the construct or concept. Delimiting, or defining the proper samples in terms of some applicable theory, is about checking whether the measurement indicators of the model constructs cover all the aspects of constructs in a relevant and representative manner (Fitzpatrick, 1983). It must be shown that the indicators are logically and theoretically related to the construct (Churchill, 1979; Pedhazur & Schmelkin, 2013). The experts' validation and the literature review are the usual criteria. Following the analysis of content, the next step is the definition of the nature of the relationships between these constructs and their measurement indicators. In other words, we must understand whether these measurement indicators reflect or define the constructs (Chin et al., 2008). Bagozzi and Fornell (1982) evoked the names of reflective construct for the first scenario and formative construct for the second scenario. The conventional graphical representation of these two types of constructs is shown in Figure 1 .

Figure 1. Conventional graphics for formative and reflective constructs



For many years, the validation of measurement models has been influenced by the classical test theory[1] across Churchill's paradigm (1979), which favors the use of the reflective measurement model (Jarvis et al., 2003). The simulation studies of Monte Carlo[2] have shown that a relationship misspecification influences the estimated values of the effects of a construct on other constructs likely to undermine the theoretical building (Jarvis et al., 2003; Bollen & Diamantopoulos, 2017). The choice between a formative and a reflective construct depends on the ontological status of the relationships between the measurement indicators and the construct to be operationalized (Bagozzi, 2011; Finn & Wang, 2014). Moreover, we must consider their conceptual differences (Bollen & Lennox, 1991; Jarvis et al., 2003). In marketing research, several examples of construct operationalization have been incorrectly specified (Jarvis et al., 2003). As an indication, it is found that the operationalization of "the E-service quality" construct (Collier & Bienstock, 2009) and of

[1] According to the classical test theory, measurement is defined according to the real score, systematic error and random error (Churchill, 1979). Its expression is given by: $X= T+S+E$, where X designates the reported score, T the real score, S the systematic error and E the random error. The real score "T" is the score that someone would have obtained under ideal conditions with a reliable and valid instrument; this is, theoretically, the perfect measurement. The systematic error "S" is attributed to the systematic deviation that occurred with the phenomenon to be measured, and the random error "E" is the source of random variances resulting from repetitive measurements. Based on this expression, we can say that variation in the latent variable causes a variation in the measurement indicators (Bollen, 1989)

[2] In order to study the variations of parameters depending on the nature of relationships between indicators and constructs

"the website quality" construct (O'cass & Carlson, 2012) is specified inappropriately. The simple comparison of the choices made in these two previous studies with the rules of Jarvis et al. (2003) reflects researchers' anxiety in handling SEM. Table 2 below presents a rule of decision to determine if the links between a construct and its indicators should be considered as formative or reflective in nature.

Table 2. Decision rules for choosing between a formative model and a reflective model.

Characteristics of the link between the construct and its indicators	**Formative**	**Reflective**
Direction of causality of the construct towards the indicators	No	Yes
✓ The indicators define the construct	✓ Yes	✓ No
✓ The indicators are the manifestations of the construct	✓ No	✓ Yes
✓ A change in the indicators affects the meaning of the construct	✓ Yes	✓ No
✓ A change in the construct affects the indicators	✓ No	✓ Yes
The indicators are mutually interchangeable	No	Yes
✓ The indicators show similar contents	✓ No	✓ Yes
✓ Removal of an indicator changes the meaning attributed to the construct	✓ Yes	✓ No
The indicators are correlated with each other	Not necessarily	Yes
✓ A change in one indicator of the construct results in a change in the other indicators of the construct	✓ No	✓ Yes
The nomological networks of the indicators differ	Yes	No
✓ The indicators should have similar antecedents and consequences	✓ No	✓ Yes

Adapted from Jarvis et al. (2003)

Step 1.2: Structural model specification

Applying SEM by analyzing the covariance structure forms part of a hypothetico-deductive scientific approach. The SEM approach requires that researchers posit a prior proposition or hypothesis concerning the causal link connecting two or more variables. Causal inferences can then be derived by evaluating the fit of the model to the data (Biddle & Marlin, 1987). Hypotheses or propositions can be derived from a general theory (for example, through induction from previous research), or they may result from attempts at an integrated explanation of prior results. Prior results can be considered the causal context. The hypotheses specify the causal links between variables. They take the multivariate nature of causality into account. Accordingly, when applying SEM, the explanatory model is formulated in a presumed context of causality and is contingent upon certain assumed causes, some of which can be imposed but also, as will be briefly explained in the next section, subject to the interpretation of statistical results.

Step 2: Model identification

The model specification (Step 1) affects the *identification of the model parameters (Step 2).* Here, we encounter Popper's concept of falsification (Yu, 2002). When the resulting equations do not yield a single solution but rather several, the model is said to be irrefutable, as it fits all data. Because there is no way to refute it, the model is useless. The same is true of exactly identified equations, which a priori seem to be the ideal state. The overriding goal, however, is to find the model that best fits the data. This is achieved by comparing the model with several alternative models. Accordingly, it is desirable for a model to be overidentified.

Ideally, this step should be performed before estimating the parameters. Indeed, Kline (2011, p. 131) states that:

> "*This exercise takes the form of a formal mathematical proof, so no actual numerical values are needed for elements of the sample covariance matrix, just symbolic representations of them. This means that model identification can—and should—be evaluated before the data are collected*".

According to (Chin et al., 2008), the meaning of the following steps is jeopardized if the researcher did not solve the problem of parameter identification. Unidentified parameters produce inconsistent estimated parameters and uninterpretable fit indices (Rigdon, 1995).

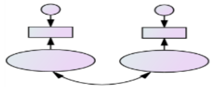
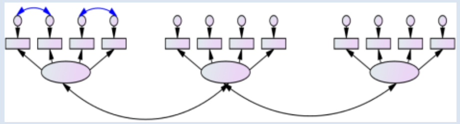
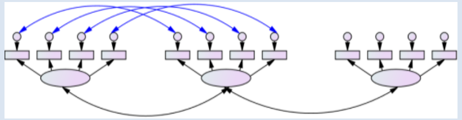
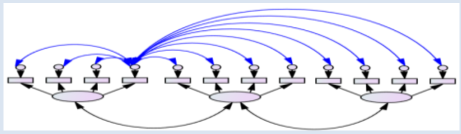
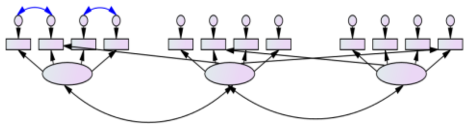
Two conditions must be met for the identification of parameters: the first is the order condition and the second is the rank condition (e.g., Kenny & Milan, 2012). The order condition states that the degree of freedom of the model must be greater than or equal to zero (Bollen, 1989). The rank condition requires algebraic determination of each model parameter whose technical difficulty depends on the model complexity (Kline, 2011). This algebraic determination can be avoided in most cases by using heuristic rules found in specialized journals (Kenny, Kashy, & Bolger, 1998). The two-step identification is normally recommended (Anderson & Gerbing, 1988). That is an identification of the measurement model followed by identification of the structural model. The merit of this approach is *ultimately* derived from the idea that focusing directly on the structural model (theoretical relation) may not be relevant as long as the measurement model is not validated (Bollen, 1989). In other words, if the sampled indicator does not measure the construct, it must be modified and/or replaced before testing the model of relations.

It is possible to classify the model recognition heuristic recommendations depending on whether it is a measurement model (reflective or formative indicator) or structural model. With regard to the reflective measurement model, several heuristic rules are recommended depending on the model complexity[3]. As an illustration, Table 3 summarizes the work of Kenny et al. (1998). For the formative measurement model, they should not be considered in isolation. Indeed, they cannot be identified because of the problems associated with the scale measurement and the presence of the error term of the latent factor (Jarvis et al., 2003). In order to be identified, it must be placed in a much broader nomological model (Diamantopoulos & Winklhofer, 2001). Table 4 presents some heuristic rules identified in the literature. The second stage consists of identifying the structural model. To identify it, Bollen and Davis (2009) state that two conditions are necessary: (1) the measurement model are identified and (2) the structural model is recursive. Otherwise, the rank condition must be met.

For a confirmatory study, the main limit to implement the above recommendations is to achieve a model that no longer complies with the originally specified theory. Thus, the researcher takes the risk of being in the uncomfortable position of having to interpret a new model and go through an exploratory study.

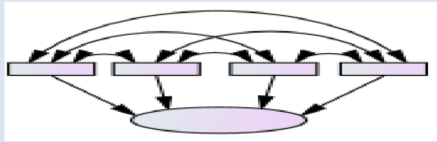
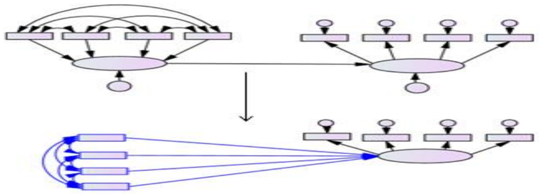
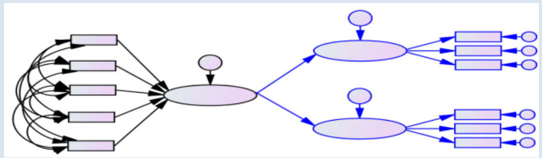
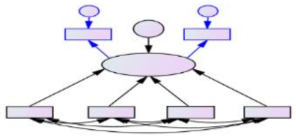
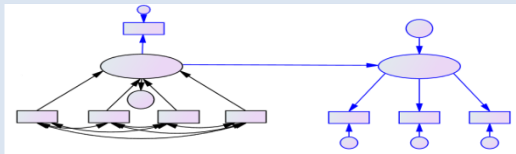
[3] A measurement model is called complex if it is multidimensional and if some indicators of one of its factors can be found on other factor(s).

Table 3. Heuristic rules of identification for a reflective model.

Models	Heuristic rules	Examples of unidentifiable models
Model A	A one-factor model is identified with at least three indicators and whose errors are not correlated.	(1) 
Model B	A model with at least two factors and at least two indicators per factor is identified. Errors should not be correlated with each other	(2) 
Model C	A reflective model with correlated errors must meet the following three conditions in order to be identified:	(3)
	➢ For each factor, one of the conditions must be verified: ○ At least the indicators' error terms are not correlated with each other ○ At least two error terms of two indicators are not correlated and; either the error terms of these two indicators are not correlated with a third error term of a different factor, or an equality constraint is enforced on the saturation factors of the two indicators	(3.1) 
	➢ For each pair of factor, there are at least two indicators, each from one of the two factors and whose errors are not correlated	(3.2) 
	➢ For each indicator, there is at least one other indicator with which it is not correlated	(3.3) 
Model D	A reflective model with at least two indicators found on other factor(s) can be identified if the following three conditions are met: ➢ Condition (3.1) is met ➢ Each pair composed of the previously selected factors must meet the condition (3.2) ➢ For each factor on which one or several indicators of other factor(s) is found, there should be at least one indicator specific to the factor alone and which is not correlated with the other indicators mentioned above (3.3)	(4) 

Source: Adapted from Kenny et al. (1998)

Table 4. Heuristic rules of identification for a formative model.

Within nomological network, one or more of these recommendations may be taken:	Examples of identifiable models
1. Enforce a zero constraint to one or more structural models or construct errors (Bollen & Davis, 2009; Jarvis et al., 2003)	
2. Decompose the formative construct if only a single structural connection is transmitted to a reflective construct (Bollen & Davis, 2009)	
3. Ensure there are at least two structural connections transmitted to two reflective constructs (Jarvis et al., 2003; MacKenzie, Podsakoff, & Jarvis, 2005)	
4. Include two reflective indicators in the formative construct (Jarvis et al., 2003; MacKenzie et al., 2005)	
5. Include a reflective indicator in the formative variable and transmit a structural connection to a reflective construct (Jarvis et al., 2003; MacKenzie et al., 2005)	

Step 3: Analysis of data structures

The points discussed in this step are part of the main criticism resulting from the work recommending the application of SEM in consumer behavior research in Marketing (Baumgartner & Homburg, 1996; Hulland et al., 1996; Martínez-López et al., 2013). Contrary to this work, the structure of our recommendations refers directly to the hypotheses underlying the application of SEM by the maximum likelihood method. These hypotheses presume (1) a large sample size, (2) independence of observations, (3) absence of missing values, and (4) a normal and continuous distribution of endogenous variables (Bollen, 1989; Kline, 2011). Compliance with these hypotheses provides unbiased, consistent and efficient standardized errors and estimators (Bollen, 1989).

Step 3.1: Sample size

Three types of recommendations to determine the minimum sample size required for the application of SEM are mainly identified in the literature. The first concerns the recommendations aiming to provide standardized rules on the sample size, regardless of the type of model to be tested. In this direction, Ding, Velicer, and Harlow (1995) suggest a sample of at least 100 to 150 observations. The second recommendation undermines the first. Indeed, some researchers believe that the sample size determination depends on several factors. Therefore, Hair et al. (2010) consider that the sample size is based (1) on the data nonnormality degree, (2) on the choice of the estimation method, (3) on the model complexity, (4) on the number of missing values and (5) on the variance of reflective construct errors. If the data distribution does not follow a normal distribution, these authors suggest collecting at least 15 observations per estimated parameter. Although the choice of estimation method solves the problem of nonnormality (*see Step 3.4: Distribution of endogenous variables*), compliance with the rule of 15 observations per estimated parameter minimizes the sampling error (Hair et al., 2010). Regarding the complexity of the model, it turns out that the more variables there are, the more it is necessary to obtain a high sample size (Hair et al., 2010). For missing values, there are several methods to handle them so as not to affect the size of the final sample. The “missing values” subsection specifically addresses the matter. Finally, regarding the error variance, a high sample size is required if the common variance is less than 0.5. Table 5 below summarizes the recommendations of Hair et al. (2010) on the sample size depending on the number of latent variables, the number of reported variables and the size of the common variance.

Table 5. Sample size according to common variance and the number of variables

Sample size	Number of latent variables	Common variance of the latent variables	Indicators per latent variable
100	Less than 5	> 0.6	More than 3
150	Less than 7	= 0.5	And measurement variables identified
300	Less than 7	>= 0.45	And/or less than 3
500	More than 7	>= 0.45 for some	And/or less than 3

Source: Hair et al. (2010)

Finally, the third recommendation deals with the identification of the minimum sample size to achieve acceptable statistical power. Based on the work of Martínez-López et al. (2013), we can see that no article of the four selected journals: Journal of Marketing Research (JMR), Journal of Consumer Research (JCR), Journal of Marketing (JM), and International Journal Research of Marketing (IJRM) has considered the power analysis. According to (Cohen, 1992, p. 155):

> *"It is not at all clear why researchers continue to ignore power analysis. The passive acceptance of this state of affairs by editors and reviewers is even more of a mystery".*

However, Martínez-López et al. (2013) reveals that approximately 30% of the studies using SEM have a low power test. This situation is worrisome, since the low power of a statistical test can lead to an erroneous research conclusion (Fan et al., 1999). Indeed, the good performance of fit indices is not sufficient to justify the non-rejection of the model (Chin, 1998). If the power of a test is low, H_0 is not rejected and the researcher might not reject a false theory inducing a type II error (Fan et al., 1999; McQuitty, 2004).

There are several methods of assessing the power of tests on SEM (e.g., MacCallum, Browne, & Sugawara, 1996; Satorra & Saris, 1985). The most recommended approach in the literature is that of MacCallum et al. (1996) (Kim, 2005; McQuitty, 2004). It derives from the method proposed by Satorra and Saris (1985) with the advantage of being able to specify in advance a bad model based on RMSEA [Root Mean Square Error of Approximation] and calculate a priori the sample size required to reach the necessary power (Kim, 2005).

To this end, MacCallum et al. (1996) suggest three (3) types of hypothesis testing: "*exact fit*" (H_0: $\varepsilon = 0$), "*close fit*" (H_0: $\varepsilon \leq a$), and "*not close fit*" (H_0: $\varepsilon \geq a$).

The "*exact fit*" test is a strong hypothesis. It amounts to testing whether the model is exactly correct. In practice, this hypothesis has little interest, given that a model can only be an approximation of reality. The use of this approach generally leads to model rejection when the sample size is high[4]. The second "*close fit*" hypothesis is, on the contrary, more realistic than the first. It makes it possible to find a more acceptable model. The third "*not close fit*" hypothesis, unlike the previous ones, corresponds to a test where the researcher's goal is to reject H_0. The advantage of this last hypothesis is to bring more validity regarding the conclusion of research. Indeed, in the first two hypotheses, if H_0 is not rejected, it does not mean that H_0 is accepted, but only that the data do not provide sufficient evidence to question it. However, in the third hypothesis, the rejection of H_0 may conclude that the model could fit the data; in other words, it may be possible that $\varepsilon < a$.

In order to calculate the minimum sample size, it is necessary to assign a priori values to an effect size π, to RMSEA index ε, and to the type I error α. These values are arbitrary. However, there are values generally accepted by the community. With respect to π, Cohen (1992) defines three types of levels: low, with π equal to 0.2, average, with π equal to 0.50, and high, with π equal to 0.80. For ε, two pairs of values are needed to calculate the power: one specific to the null hypothesis H_0, that is ε_0, and one specific to the alternative hypothesis H_a, that is ε_a. Four types of interpretation arise therefrom: "*exact fit*" for $\varepsilon_0 = 0$ and $\varepsilon_a = 0.05$, "*close fit*" for $\varepsilon_0 = 0.05$ and $\varepsilon_a = 0.08$, and "*not close fit*" for $\varepsilon_0 = 0.05$ and $\varepsilon_a = 0.01$ (MacCallum et al., 1996). Finally, for the alpha value, it is usual to choose between 1%, 5% and 10%. With regard to the method applications, the SAS [Statistical Analysis System] code lines can be found in the article by MacCallum et al. (1996) .

Finally, the first and third recommendations are suggested to allow a priori diagnostics of the sample characteristic (e.g. normality test). The sample size to be used is the highest one of these two recommendations. The second recommendation should be followed after performing the diagnostics. Finally, remember that it is important that the sample, as in any sampling procedure, be representative of the population[5].

[4] Yet, it is a necessary condition to obtain unbiased and asymptotically convergent parameters.

[5] Statistically, the representative word is used here as a misnomer.

Step 3.2: Independence of data

Several factors may be at the origin of data dependency. We can include mainly the paired data and multilevel phenomena. While the characteristics of paired data are easy to diagnose, they are less obvious for multilevel phenomena. No warning on this topic has been put forward in the SEM applications journals regarding the field of this study (Baumgartner & Homburg, 1996; Chin et al., 2008; Hulland et al., 1996; Martínez-López et al., 2013; Richter, Sinkovics, Ringle, & Schlägel, 2016). This situation may be explained by the fact that the multilevel approach pertains to an advanced approach of the latent variable SEM[6]. However, we must admit the urgency of treating it systematically. Indeed, it has been accepted that considering a multilevel phenomenon only at the individual level leads to inefficient estimators and standardized underestimated errors (Hox, Moerbeek, & Van De Schoot, 2010. According to Curran (2003, p. 534):

> "*[…], failure to explicitly model the nested structure of the data may significantly preclude our ability to test certain questions of interest*".

Sempé (2000) comparison of RMSEA fit indices of CFA models belonging to single-level social circles with the multi-level ones illustrates this importance. Indeed, if for the single-level CFA, the probability for RMSEA of being below 0.05 is 0%, it is slightly less than 100% for the multi-level model. In this context, the multi-level model has significantly reduced the unexplained error. For further information on the approaches of multi-level phenomena modeling, the work of Hox, Moerbeek, and van de Schoot (2010) can be consulted.

Step 3.3: Non-responses

Regarding the missing values, it seems that only 9.6% of the articles in the marketing field discuss the numbers and treatments of missing data (Hulland et al., 1996). The treatment of missing values does not seem to hold the attention of other authors in SEM application recommendations. The aim of these treatments, however, is to obtain (1) non-biased parameters, (2) a good evaluation of the variance of estimated parameters, and (3) high power (Graham & Coffman, 2012). According to Kline (2011), the proportion of missing values for a variable less than 5% has a relatively low impact if it is due to nonsystematic causes. Several methods can be used to reduce non-responses, especially when designing an optimal questionnaire. Researchers should consider the

[6] See Marcoulides and Schumacker (1996).

questionnaire’s completion time, clarity, legibility, and legal obligations, along with the procedures for re-approaching a subject for follow-up questions. To identify any biases that may be created by non-responses, non-respondents can be approached anew to determine why they failed to respond and correct any resulting biases, using suitable statistical techniques.

Rubin (1976) distinguishes three non-response mechanisms: (i) responses missing completely at random (MCAR), (ii) responses missing at random (MAR), and (iii) responses not missing at random (NMAR).

The MCAR mechanism means that the non-response process is a random event independent of the variables being studied. For example, say that a survey seeks to assess students' perceptions of the quality with which a subject is taught. If non-responses are caused by the absence of certain students who are ill, then the grounds for their illness is completely independent of the study variable "quality of teaching." In this case, the estimators are not biased. The only statistic that can be affected is the power of the test, if there are many absences on these grounds.

For the MAR mechanism, the mechanism underlying non-responses can be determined among the different variables of the survey dataset. In this case, non-responses can be modeled by including the cause variable(s) that are thought to predict non-response. For example, non-responses to a survey on contraception methods can be influenced by the culture of the community to which the observed individual belongs, e.g., a religious community, as the subject matter of the survey may embarrass the respondent. In this case, a predictable factor causing non-response is belonging to a certain sort of community. If this kind of situation is not corrected by allowing for including the variables causing non-response, a bias can arise for the estimators (Collins, Schafer, & Kam, 2001).

Lastly, for the MNAR mechanism, an unknown characteristic of non-response is behind the very act of participating in the survey. This is, for example, the typical case of income surveys, where individuals with large incomes are reluctant to provide information on this subject. This kind of situation also creates biases for estimators (Graham & Coffman, 2012). Nonetheless, the biases generated are not as great as those caused by using traditional methods (Collins et al., 2001).

To address non-responses, researchers must know the mechanism underlying the non-responses. Unfortunately, no specific rule exists for deciding on a particular mechanism. The answer lies in researchers' knowledge of their field of research and which hypothesis they judge to be pertinent regarding the non-response mechanism applicable to their survey results. Statistical tools are available for testing whether the missing data are distributed according to the MCAR or the MAR pattern, such as Little's MCAR test. For the MAR mechanism, researchers can test for correlation between the variables thought to cause non-responses and the study variables. Nonetheless, these statistical tests merely reinforce the hypothesis that the researcher initially accepted. They do not constitute confirmation or identification of the mechanism explaining the distribution of non-responses.

To process non-responses, the literature distinguishes between two principal types of processing: traditional processing methods such as paired exclusion or list exclusion, and modern processing by the maximum likelihood method and the multiple imputation method. In general, exclusion in pairs is discouraged because it creates problems of variability depending on sample size, thus undermining the premise that the covariance structure follows a multi-normal distribution (Browne, 1994). It is liable to bias fit indices based on Chi-squared statistics (Marsh, Balla, & McDonald, 1988). Parameters can also be significantly biased, preventing direct estimation of standard errors (Graham & Coffman, 2012). Finally, list exclusion is unsuitable whenever the sample size is small because it generates biased parameters and standard errors when the share of missing values is large (Enders & Bandalos, 2001).

With respect to modern methods for addressing non-responses, methods such as FIML and the multiple imputation method solve the problem of statistical power under the assumption of an MCAR mechanism. In the case of an MAR assumption, FIML provides corrected standard errors and a corrected Chi-squared test result, provided that the premises for its application are fulfilled. When the assumptions underlying FIML are not fulfilled, it is preferable to use the multiple imputation method instead, using a combination of auxiliary variables that can significantly improve statistical power. For estimators, if the auxiliary variables are omitted from the non-response analysis (whether for FIML or for multiple imputation), the degree of bias depends on two things: the correlation between non-response and the study variables (the higher the correlation, the greater the bias), as well as the percentage of non-responses in the survey (the more non-responses there are, the greater the bias) (Collins et al., 2001). For example, for a non-response rate of 50%, if the correlation between a study variable and the variable thought to be its cause is 0.9, then the simulation results indicate significant biases for the estimators. By contrast, for a correlation of 0.4 between the two variables and a non-response rate of 50%, the bias is relatively small.

Currently, publishers of off-the-shelf statistical software, such as LISREL, SAS, and SPSS, sell easy-to-use programs that address non-responses due to MCAR and MAR. By contrast, both detecting and correcting an MNAR mechanism are challenging. According to Arbuckle (1996), methods for processing missing values are valid only for randomly missing data (MCAR and MAR). Collins et al. (2001) suggest that modern approaches are nonetheless able to attenuate some of the biases resulting from MNAR. Overall, the MNAR mechanism requires sophisticated modeling of a sort that is not routinely available in off-the-shelf software (e.g., Matei and Ranalli, 2015).

Step 3.4: Distribution of endogenous variables

Regarding the assumption of normality of endogenous variables, a high degree of nonnormality impacts the estimated parameters and standardized errors that become underestimated (DiStefano, 2002; Kline, 2011). This situation usually leads to the rejection of hypotheses resulting from the increased likelihood of making a type I error. As there are no criteria leading to a common agreement in defining the degree of normality of variables, those developed by Curran, West, and Finch (1996) may be used. For *Skewness* ~ 2 and *Kurtosis* ~ 7, the degree of nonnormality is considered moderate. It is considered high when Skewness ~ *3* and *Kurtosis* ~ 21.

In the social sciences, the problem of normality is generally inevitable (Harlow, 1985). Unfortunately, it appears that only 6.5% of the journals collected by Martínez-López et al. (2013) have studied the distribution of their variables. For the rest, the assumption of data normality may have been admitted without performing statistical tests. In addition, in consumer behavior research in marketing, most measurement variables are categorical (Bearden, Netemeyer, & Haws, 2011). The choice of treating the categorical dependent variable as a continuous variable can be one of the reasons for violation of the assumption of normality (Bollen, 1989). The problem with categorical variables is that the Pearson correlation matrix tends to underestimate the degree of association, thereby resulting in an underestimation of parameters (DiStefano, 2002; Kline, 2011). For approximately normal variables with at least five categories, the estimated parameters are slightly biased (Babakus, Ferguson Jr, & Jöreskog, 1987; Muthén & Kaplan, 1985; West, Finch, & Curran, 1995). For the variables with high *Kurtosis* and *Skewness* degrees, the level of bias becomes high (Dolan, 1994; West et al., 1995). Moreover, the standardized errors are underestimated, and can cause a type I error (Dolan, 1994; Muthén & Kaplan, 1985; West et al., 1995). A similar situation occurs for variables under five categories (Dolan, 1994).

The maximum likelihood method[7] (*ML*) remains the estimation method mostly used in the consumer behavior study in marketing (Baumgartner & Homburg, 1996; Martínez-López et al., 2013). The work of Martínez-López et al. (2013) reveals that 81% of the articles did not mention the method used, leaving doubts about the research conclusions.

We can classify the solutions addressing the data nonnormality problems into two categories: the transformation of variables and estimation methods to get around the assumption of normality. The first method relates to a monotonic transformation of the data distribution. Several types of transformation function may be used depending on the nature of the flattening degree and/or on the data distribution asymmetry. For example, the Box and Cox (1982) methods to transform nonnormal variables. The disadvantages of the transformation method lie in the interpretation of scales of transformed data and its application limited to continuous variables.

Regarding the second method, it can be distinguished according to whether the variables are categorical or continuous. For continuous variables, we can mention:

1- *The estimation methods by generalized least squares (GLS) and by maximum likelihood (ML).*

   GLS better overcomes the violation of assumption of normality if the sample size is small (Olsson et al., 2000). However, it is recommended only when the model is simple. Indeed, it is impracticable for complex models, and can ultimately lead to very unsatisfactory results (Hu & Bentler, 1995). For the maximum likelihood method, this procedure is robust if the variables' degree of nonnormality is low (Chou & Bentler, 1995). When the degree of nonnormality increases, the method tends to underestimate statistics such as standard error and fit indices, including TLI and CFI (West et al., 1995). Despites these previous results, it appears that ML is more insensitive than GLS to sample size and kurtosis (Olsson et al., 2000). ML was also found to be more stable and more accurate on parameters and fit indices estimates (Olsson et al., 2000).

[7] The minimization function of the *ML* method is given by the following formula: $F = \frac{1}{2} tr[([S-\sum(\theta)]W^{-1})^2]$, where S the empirical variance-covariance matrix, $\sum(\theta)$ the implied variance-covariance matrix from the postulated model *W* the weighting matrix $\sum(\theta)^{-1}$.

2- *The "Asymptotically Distribution-Free (ADF)" estimation method*. Browne (1984) proposes the *ADF*[8] estimation method for unknown data distributions. On the contrary, this method requires a sample size with at least 1,000 observations (West et al., 1995). The sample size increases according to the model complexity and/or degree of nonnormality of the model variable (Hu, Bentler, & Kano, 1992). In consumer behavior research in Marketing, large samples are rarely used if we refer to the average size of samples in the empirical journals on *SEM* application. It varies between 178 and 259 observations (Baumgartner & Homburg, 1996; Martínez-López et al., 2013).

3- *The corrected standard error of Satorra-Bentler (S-B).* The corrected standard error of *S-B* is generally used with the ML method. The correction made does not apply to parameters to be estimated, because the latter is not biased (Curran et al., 1996). It impacts the standard error and the Chi-squared that are biased in the case of the *ML* method under an increasing nonnormality (Chou, Bentler, & Satorra, 1991; Curran et al., 1996). Chou and Bentler (1995) found that the results of this method proved effective in relation to the estimation methods *ML* and *ADF*. In addition, it appears that this method is preferable to the *ADF* estimation method when the sample size is not too large (Chou et al., 1991).

For the categorical variables, we can mention:

1- *The Maximum Likelihood (ML) estimation method.* A comparison of performance between the *ML* methods by the Pearson correlation matrix and by the polychoric/polyserial matrix shows that the latter provides the best unbiased and consistent estimators (Brown, 1989). On the contrary, the studies conducted by Dolan (1994) revealed unbiased and consistent estimators for variables with symmetrical distribution with more than seven categories. Others, such as Flora and Curran (2004), have also found similar results. However, for variables with high degrees of nonnormality[9], the *ML* method with polychoric matrices proves to be limited, since it provides biased estimators (Flora & Curran, 2004). For the standard error, it seems to be overestimated (Lei, 2009)

[8] The minimization function of the ADF estimation method is given by the following formula: $F = (s-\sigma)' W^{-1}(s-\sigma)$, where s represents the elements of the empirical variance-covariance matrix and σ of the theoretical variance-covariance matrix. The weighting matrix W is the asymptotic variance-covariance matrix from the empirical variance-covariance matrix. The elements of this matrix are obtained using the empirical covariance matrix with the 4th order moment : $W_{ij,kl} = s_{i,j,k,l} - s_{ijskl}$ (Browne, 1984). The problem of this method from the technical point of view is the inverse calculation of the W matrix $^{-1}$ which is cumbersome to calculate (a matrix of size: ½ (p+q) (p+q+1), where p is the number of exogenous variables reported and q the number of endogenous variables reported (Bollen, 1989), requiring a large sample to achieve convergence.

[9] This is the degree of nonnormality of the underlying latent variable.

2- *The ADF estimation method*. The application of the *ADF*[10] method provides slightly biased estimations, whether the variables are dichotomous (Muthén & Kaplan, 1985) or categorical (with *Skewness* = 2.5 and/or Kurtosis = 6) (DiStefano, 2002). However, for the standard error, it seems to be underestimated and decreases as the size decreases or as the size of the model and/or the degree of normality increases (DiStefano, 2002).

3- *The corrected standard error of Satorra-Bentler (S-B).* The correction relates only to the standard error and not to the parameter to be estimated. The correction is much more accurate for Skewness ~ 2 and/or Kurtosis ~ 6, even for categorical variables up to three categories (DiStefano, 2002). For parameters, regarding the categorical type variables for those under 5 categories, the application of the maximum likelihood standard method induces slightly underestimated parameters (Babakus et al., 1987).

4- *The robust methods "*Weighted Least Squares *Mean-adjusted* (*WLSM)" and "*Weighted Least Squares Mean- and Variance-adjusted (*WLSMV*)*".* Muthen (1983) suggests two robust methods *WLSM* and *WLSMV*[11] as an alternative to the *ADF* of Browne (1984). It appears that these methods are relatively more efficient than the *ADF* method (Flora & Curran, 2004). In addition, the results of Lei's studies (2009) revealed that for a sample size of at least 250 observations, the *S-B* and *WLSMV* methods show the same performance. On the contrary, *WLSMV* is less efficient compared to the *S-B* method for a sample size below 100 (Lei, 2009).

In summary, the *ADF* method is not recommended, whether for continuous or categorical variables. For approximately normal variables, the ML estimation method can be used in both types of variables, provided that the number of classes is at least five for categorical variables and the corrected standard error of *S-B* is used for continuous variables. If the categories are less than five, it is recommended to use the *S-B* method to estimate the standard error. For variables with relatively high nonnormality degrees for continuous variables, the *S-B* method seems to be a satisfactory

---

[10] The *ADF* estimation method with categorical variables is slightly different from the *ADF* of continuous variables. This method uses an approach that transforms the discrete variables into a continuous latent variable. For example, for a variable with three categories, the transformation is defined as follows: $y = c$ if (if $\pi c < y^* \leq \pi c+1$), where $\pi c = \Phi^{-1}[\sum_{k=1}^{c} N_k / N]$, k= 1, 2, 3 with $\Phi^{-1}$the inverse function of the distribution of a normal law, N_k the number of observations in the category k and N the size of the total sample. The minimization function is the same: $F = (r-\rho)' W^{-1}(r-\rho)$, but the correlation matrix used is the polychoric/polyserial matrix

[11] Both methods use the asymptotic variance-covariance matrix for weighting, just like the *ADF* method. The advantage of these methods is the use of matrices with complete information that do not require inverse calculation for standard error. Furthermore, it uses only the diagonal of the matrix to estimate the parameters. Therefore, the sample size to find a consistent estimator is lower and more acceptable than the *ADF method.* Currently, these estimation methods are available for the *Mplus* software.

solution. Finally, for categorical variables, the robust methods *WLSM* and *WLSMV* are recommended.[12]

Step 4: Empirical validation of psychometric qualities of measurement variables

If, in the first step, we discussed the ontological considerations of the nature of relation of the reported variables and latent factor, this *Step 4* focuses more on the empirical validation of these relations through constructs validity criteria. Indeed, this procedure requires a fairly substantial collection, analysis and processing of data in order to achieve a quantitative analysis. Based on the data, it is about assessing the degree to which a measurement determines the construct it claims to assess (Bernstein & Nunnally, 1994). The trait validity and nomological validity are the criteria mostly used to support the construct validity (Campbell, 1960). The trait validity refers to the investigative efforts of the measurement reliability, convergent validity and discriminant validity (Peter, 1981). Reliability is a necessary condition for a valid measurement instrument. A reliable instrument must provide identical responses every time it is used (Thiétart, 2014). Convergent validity requires the measurement of a construct to be perfectly correlated with the other measurements of the same construct. On the contrary, discriminant validity assesses the degree to which the measurements of the same construct differ from the measurements of other constructs (Peter, 1981). However, this psychometric quality validation process depends on the nature of the measurement variables, whether it is a reflective model or a formative model (MacKenzie et al., 2005).

The most commonly used reliability test for reflective measurement in the field of consumer behavior research in Marketing is Cronbach's alpha (Martínez-López et al., 2013). It appears, however, that this coefficient has a certain variability based on the number of items included in the test (Garver & Mentzer, 1999). Furthermore, all items are equally weighted, which normally is not always the case (Bollen, 1989). Jöreskog's rho coefficient rarely used in the journals rather responds to the characteristics of structural equations, since it explicitly incorporates the terms of the error and the factorial contribution of measurement indicators (Jöreskog, 1969). Therefore, this coefficient may replace Cronbach's alpha that will always be used as reference.

[12] We have seen that the *S-B* method provides better results than those obtained by the *WLSM* and *WLSMV* methods for sample sizes below 100. We believe it is not recommended, considering that the SEM application standard is to work with samples with at least 100 observations (ref. Subsection: sample size).

Regarding the convergent validity and the discriminant validity, if the exploratory factor analysis (EFA) and the method of Multitrait-Multimethod Matrix (MTMM) of Campbell and Fiske (1959) have been traditionally used to establish the convergent and discriminant validity (Laurent & Kapferer, 1985; Zaichkowsky, 1985), the confirmatory factor analysis[13] (*CFA*) is more suitable to the confirmatory approach, because the factor structure is defined a priori (O'Leary-Kelly & Vokurka, 1998). Indeed, the exploratory method rather explores factor structures. In general, we find different results depending on the method used (Gerbing & Anderson, 1988). As regards the MTMM matrix, it is not recommended due to the difficulty of selecting the method to deploy (Kalleberg & Kluegel, 1975), but also due to the convergence problems (Lomax & Schumacker, 2012). Another alternative, the "*Correlated Uniqueness Model*" is recommended, as it gives better results (Lomax & Schumacker, 2012). This method is tenable insofar as the measurement variance is distributed between the trait, method and error variance (Bollen, 1989).

For variables of formative measurement, the reliability test in internal consistency does not make sense. Indeed, as the indicators are not reflections of the construct, the correlations between indicators can be zero or negative (Diamantopoulos & Winklhofer, 2001). In this same logic, the test of convergent validity is also not relevant (Bollen, 1989; MacKenzie et al., 2005)

The reliability of a formative variable is evaluated mainly from the degree of multicollinearity (Diamantopoulos & Winklhofer, 2001), of the significance test of coefficients (MacKenzie et al., 2005) and; optionally, of the test-retest method (Petter, Straub, & Rai, 2007). On the theoretical level, the presence of multicollinearity of data means that one or more indicators are not a separate facet of the construct. The elimination of formative indicators of the measurement model is therefore encouraged (Diamantopoulos & Winklhofer, 2001; MacKenzie et al., 2005). In addition, a high degree of multicollinearity leads to biased estimates of parameters and does not make it

[13] In *CFA*, convergent validity is normally obtained by examining the adequacy of model measurement to the data, direction and statistical significance of the estimated parameters between the latent constructs and their reflective indicators (Fornell & Larcker, 1981). The calculation of the average variance derived for a particular construct is often used to assess the convergent validity (Fornell & Larcker, 1981). The convergent validity hypothesis is satisfactory for the confirmatory study if the fit indices do not reject the measurement model. In addition, the weights of indicators should be greater than 0.707 to prove that more than half of the variance is captured by the latent construct (Chin, 1998; Gefen, Straub, & Boudreau, 2000). As part of *CFA*, discriminant validity is typically evaluated by using two methods. The first method is to calculate the variance shared between the constructs and to check if it is below the variances extracted on the constructs concerned (Fornell & Larcker, 1981). Thus, a relatively low correlation between a pair of latent variables is the first indication of the presence of discriminant validity. The absolute maximum should not typically exceed $r = 0.707$ to match a shared variance that is not greater than the recommended minimum of 50% of average variance extracted (Fornell & Larcker, 1981). Then, it is necessary to compare $\chi 2$ of the original model with an alternative model in which the constructs concerned are united as a one-dimensional construct. If $\chi 2$ is substantially smaller in the original model, the discriminant validity is proven (Gefen et al., 2000). The second method to assess the discriminant validity is to examine all possible pairs of construct in a series of two CFA model factors (Bagozzi & Yi, 1989). Specifically, each *CFA* model of the two halves is performed twice in a first step, binding the correlation coefficient "Φ" between two variables to the unit, and in a second step, allowing "Φ" to vary freely. Based on the results of different Chi-square tests (Anderson & Gerbing, 1988), we can ensure the discriminant validity between the measurements by showing that the non-binding model provides more performance than the associated binding model when "Φ" is equal to one (1).

possible to isolate the particular influence of an indicator in the measurement model (Diamantopoulos & Winklhofer, 2001; MacKenzie et al., 2005). On the contrary, their elimination should not be separated from conceptual considerations (Diamantopoulos & Winklhofer, 2001). In order to detect the presence of multicollinearity, the analysis of the correlation matrix and/or tolerance of indicators and/or the variance inflation factor are recommended (Diamantopoulos & Siguaw, 2006).

For the discriminant validity, the principle used for a reflective measurement is applicable; the common variance between two constructs should be less than 50% (MacKenzie, Podsakoff, & Podsakoff, 2011) and the coefficients of indicators defining the constructs must be significant and in the direction postulated by theory (MacKenzie et al., 2005).

Finally, for the nomological validity, for either formative or reflective constructs, it should be tested if the predictions based on a concept that an instrument is supposed to be measured are confirmed (MacKenzie et al., 2011). For example, we can compare the degree of similarity between the results obtained on the relation between measurements and those of past work

Step 5: Evaluation of fit of the model

The fit of a structural model can be evaluated through fit indices and structural coefficients. As stated by Olsson et al. (2000, p. 564):

> "*the better the empirical fit, and the more statistically significant the parameter estimates, the more faith one has in theoretical model*."

Step 5.1: Evaluation through fit indices

In this substep, the researcher evaluates the difference between the empirical and theoretical variance-covariance matrices. Under a confirmatory approach, the model is rejected[14] if the difference is significant. In the case of an exploratory approach, the model will be re-specified (see Step 6). However, no consensus could be established on the choice of existing indices to assess the fit of a model. The complexity of choice primarily arises from two situations: (1) several fit indices have been developed for different objectives and/or with different conditions complicating their

[14] This means that the data collected do not fit the postulated model.

comparisons; and (2) with the exception of a few index minorities[15], the distributions of most indices are not known so as to conclude statistically if a difference is significant or not.

According to Gerbing and Anderson (1992), it is possible to classify the indices into three categories, as follows: overall fit index, model comparison index and parsimony index[16]. These three categories of indices complement each other to interpret the quality of fit of the model. Choosing among these categories of indices is like, according to (Gerbing & Anderson, 1992), comparing its complexity to answering the question "what is the best car on the market ?". The best car on the market depends on the purpose or preference of everyone. Indeed, some would favor performance, others the design, and others still the safety.

Several studies have shown that several factors such as sample size, degree of model complexity, degree of nonnormality, estimation methods and number of indicators variables affect the results of the fit indices (e.g., Fan et al., 1999; Gerbing & Anderson, 1993; Sharma, Mukherjee, Kumar, & Dillon, 2005). This situation made researchers question the characteristics of a good fit indices and the threshold values (Tanaka, 1993). In this context, Breckler (1990) recommends selecting indices able to better assess the adequacy of the model regarding the data instead of using the indices that validate the model. The standard fit indices recommended by (Hu & Bentler, 1999) are the most widely used in the literature. Among existing decision rules, the following combinations are the most commonly used in the literature: TLI $\geq$ 0.96 and SRMR $\leq$ 0.09, RMSEA $\leq$ 0.06 and SRMR $\leq$ 0.09, CFI $\geq$ 0.95 and SRMR $\leq$ 0.09. The benefits and drawbacks of these different measures are listed in Table 6. These indices are used because of their low sensitivity to the sample size and their ability to detect a model misspecification in comparison to the performance of other indices. Some researchers do not advocate using fit indices, however, instead preferring predictive interpretations of the model. Kline (2011), citing Barrett (2007), advocates the use of R^2 as a proxy for the model's predictive power or for decomposing the effect of the dependent variable.

Step 5.2: Interpretation of the structural coefficients

The interpretations of structural coefficients are only relevant if the selected fit indices do not reject the model to be tested. Indeed, fit indices measure how well the estimated parameters are able to

[15] For example, χ^2 and *RMSEA*.

[16] The absolute indices measure the fit of the global model. With regard to the incremental indices, they are intended to compare two or more models. The tested model should be compared with either a null model and/or alternative models. For parsimony indices, the objective is to find a model with the least possible parameter by evaluating the parameters that would only bring a marginal gain to the fit.

match the sample covariance. In this step, the researcher checks if the structural coefficients - estimated parameters - correspond to the theory in terms of meaning of the relation. That is to say, are the structural coefficients statistically significant as expected and do the sign of the coefficients indicate the same direction as the theoretical proposal? However, statistically significant coefficient is not sufficient to have meaningful interpretation. According to Chin (1998), standardized path should be at least 0.2 (ideally above 0.30) in order to be considered meaningful for discussion.

Table 6. Advantages and disadvantages of the recommended fit indices

Indices	Advantages	Disadvantages
χ^2	➢ Known distribution. Possible statistical tests. ➢ Historical reference	➢ Under a severe nonnormality condition, the test tends to reject a real model more frequently (McIntosh, 2007) ➢ The test tends to reject the models when the sample size is large (Jöreskog & Sörbom, 1993) ➢ The power of the test is low when the sample size is small (Kenny & McCoach, 2003)
RMSEA	➢ Favors the parsimonious model (Diamantopoulos & Siguaw, 2000) ➢ Known distribution. Possible statistical tests (MacCallum et al., 1996; McQuitty, 2004). ➢ For CFA, under the assumption of normality, the more indicators there are, the more efficient the index is (Kenny & McCoach, 2003) ➢ Sensitive to the model misspecification (Fan et al., 1999) ➢ Less influenced by the estimation method (Fan et al., 1999)	➢ The index tends to frequently reject a correctly specified model with a small sample size ($N < 250$). The goodness of fit deteriorates when the number of variables increases (Kenny & McCoach, 2003)
CFI	➢ Index adjusted to the complexity and parsimony for models to be compared (Iacobucci, 2010) ➢ Index less sensitive to the sample size (Fan et al., 1999) ➢ Better performance in terms of statistical power and robustness (Hu & Bentler, 1998)	➢ Under the assumption of normality, the index is less efficient with a model with one/two factors than with a large number of indicators (Kenny & McCoach, 2003)
TLI	➢ Index less sensitive to the sample size (Hu & Bentler, 1995)	➢ Under the assumption of normality, it is less efficient for a proper model with several variables (Kenny & McCoach, 2003)
SRMR	➢ The index is efficient provided that the saturation coefficients are high (Anderson & Gerbing, 1984) ➢ Less sensitive to the violation of the assumption of normality	

Step 6: Model respecification

A model requires a change only (1) if it fits the data but is overparameterized; or (2) if the statistical tests and/or fit indices conclude with the model rejection (Hu & Bentler, 1998). With a view to confirmatory procedure, a subsequent change amounts to making an exploratory or development approach of the model (Martínez-López et al., 2013). This approach is the subject of several controversies on the question of validity of the model (MacCallum & Austin, 2000). According to Brannick (1995), finding the correct model based on this approach is very likely a matter of luck. For Hughes, Price, and Marrs (1986), a model development strategy can be adopted subject to prudent and rigorous approaches. In this sense, three recommendations from the literature can be suggested: (1) the respecification should be based on theoretical considerations for the case of a change in structural model, and on the validity of content for the measurement variable (Gerbing & Anderson, 1988); (2) the results derived from the model respecification must be replicated to another independent sample to justify its generalization (MacCallum, 1995); (3) the inherent limit to the specific context of the study should be mentioned (Bollen, 1989). Another point that is also worth making after respecifying the model is to subsequently evaluate the power of a test. It means to check if the required statistical power is still met with the new model. Indeed, in the possible event of increased degree of freedom of the model, the sample size subsequently determined may be insufficient. In such a case, a new determination of the required minimum size is performed by returning to step 3 "*Analysis of the data structure*". In the case of a decrease, it is not necessary to re-evaluate the statistical power, since it is an increasing function of the number of the degree of freedom for a determined sample size *n*.

We can identify two major types of model respecification methods: "*backward search*" and "*forward search*" (MacCallum, 1990). The "*backward search*" results in the restriction of free parameter(s) and/or deletion thereof in order to obtain a parsimonious model. In this case, the degree of freedom and the Chi-square increase. On the contrary, the "*forward search*", consists of releasing the constraints of some parameters set a priori for a better fit to the data. Therefore, the degree of freedom and the Chi-square decrease.

The study results of Chou and Bentler (1990) have shown that the three tests: the modification indices (*MI*), the likelihood ratio (*LR*), and the Wald test (*W*) show almost similar performance for a fairly reasonable sample size. The *MI* test tends to suggest more parameters than necessary when the sample size is large. This situation can be corrected by completing the *MI* method with the Wald test. Other additional tools are available to identify a model misspecification. This is the significance test by the Student *t* test and *EPC* (*Expected Change Statistics*) (Saris, Satorra, & Van der Veld, 2009). *EPC* was suggested by Saris, Satorra, and Sörbom (1987) to select parameters to be added to models in order to improve the goodness of fit, but also to identify a model misspecification. The fixed parameter associated with the highest value of *EPC* indicates the worst model specification. This implies that this parameter could be estimated freely. Unfortunately, there is no consensus on an appreciation of greatness of *EPC*. For example, Kaplan (1989) considers that a value greater than 0.10 is considered a sign of a relation misspecification between two constructs. Although there are no threshold values for assessing the *EPC*, the most important is that the respecification must be validated by theory (Whittaker, 2012). The simulation study to compare the performance of a two-factor model depending on the sample size, the value of the saturation coefficient, the correlation of variables revealed a good performance of the two methods. However, *EPC* is relatively efficient compared to *MI*.

Table 7 below summarizes the synthesis of Saris et al. (1987), making it possible to interpret the *MI* and *EPC* results.

Table 7. Interpretation guide for MI and EPC results

Results found for a parameter	**Recommended procedures**
If high *MI* and high *EPC*	It is recommended to estimate the parameter freely
If high *MI* and low *EPC*	It is not recommended to estimate the parameter freely
If low *MI* and high *EPC*	An evaluation of the statistical power is recommended before deciding whether or not to estimate the parameter freely
If low *MI* and low *EPC*	It is not recommended to estimate the parameter freely

Source: Saris et al. (1987)

Step 7: Generalization of the theoretical model

The generalization of a model is its ability to match the data regularly (Preacher, 2006). Researchers wish to ascertain not only whether their models fit their own data but also how well they fit outside data (Preacher, 2006). Conversely, one or more additional models might also fit the same sorts of data, as occurs with equivalent models.

Step 7.1: Replication of results on a new sample.

> *"Many, if not most, SEM applications involve a single, one-shot model that may have had post hoc adjustments. This provides little or no opportunity to assemble causal evidence." (Bullock et al., 1994, p. 260)*

The validation of a theoretical model refers to its ability to generalize itself (Martínez-López et al., 2013). Bagozzi and Yi (1988) state that it is possible to generalize a result on a sample resulting from a population if an independent sample drawn from the same population gives the same results. Indeed, the reproducibility of results is needed to reduce uncertainty in the results derived from SEM (Steiger, 1990). Access to new samples, however, is problematic. An alternative method that could be used is the application of the *Cross Validation Index (CVI)* (Cudeck & Browne, 1983). Several methods of cross validation exist. Here, we only present two such methods. The first method is based on a Chi-squared difference test, while the second is that advocated by Cudeck and Browne (1983).

The cross-validation method for the Chi-squared difference test permits the evaluation of the invariance of the model's results by using various samples from the same population. The method is executed within the framework of a hypothesis-testing approach. The generalization is not rejected if the null hypothesis is not rejected. The hypothesis tested is the following:

> *H0: The parameters of the measurement model being tested (i.e., its weighting factors, factor variances and covariances, and measurement errors) are invariant on two samples from the same population.*
>
> *Against*
>
> *H1: At least one of the parameters of the measurement model is not identical on two (2) samples.*

The method of Cudeck and Browne (1983) consists of dividing the sample into two sub-samples: a calibration sample and a validation sample. From the calibration sample, we seek to obtain the implicit variance-covariance structure of the Σ_a model. Then, we evaluate the difference between the Σ_a model and the empirical variance-covariance matrix of the validation sample S_b. The expression CVI is defined as follows:

$$CVI = F(S_b, \Sigma_a)$$

CVI near zero indicates that the model is validated. In the event of a model comparison, we use the model with the lowest CVI. The advantage of this method is that it is available under *LISREL*.

Step 7.2: Equivalent model

Although the issue on the threshold of acceptability of data fits of indices is widely debated in the literature (e.g., Bandalos & Finney, 2010; Hu & Bentler, 1999), this does not mean in any case that it is possible to confirm a model, even if a consensus was found (MacCallum, Wegener, Uchino, & Fabrigar, 1993). Indeed, work on equivalent models has shown that it is possible that several types of models using the same variables may have the same covariance matrix (Williams, 2012).

These equivalent models may be derived from the data structure of the selected sample and/or from the relation structure inherent to the specified model (Lee & Hershberger, 1990). This last point is problematic, because we must choose from among several types of potentially relevant models. According to MacCallum et al. (1993), this problem is one of the main SEM limitations to validate a theory. Indeed, if the conclusion of a research is based on the interpretation of the sign, of the significance test and extent of the estimated parameters, several studies have demonstrated that these results vary considerably in different equivalent models (e.g., Breckler, 1990; MacCallum et al., 1993).

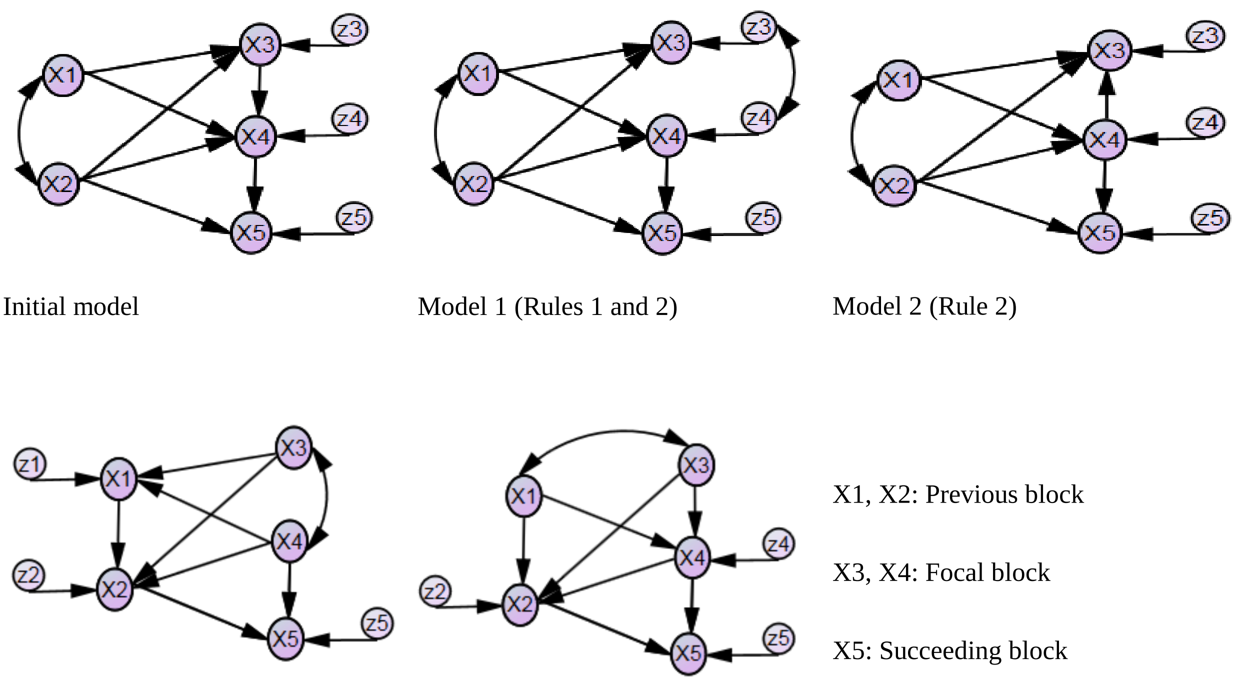
In order to create equivalent models, Lee and Hershberger (1990) and Hershberger (2006) suggest a method called "*replacement rules*". The application of these rules requires partitioning of the original model to be changed into three parts: the previous block, the focal block and the succeeding block, and change according to the replacement rules of the nature of relations between the variables of the focal block (Lee & Hershberger, 1990; MacCallum et al., 1993; Stelzl, 1986).

Provided that the relation between the blocks and within the focal block are recursive, the replacement rules state that:

Each variable of the focal block must have or include the same background belonging to the previous block. Under these conditions, the following rules may apply: the direction of causality between the variables X and Y in the focal block can be substituted by an association link (Y correlated to X). If the two variables of the focal block have the same background in the previous block, the direction of causality between the variables X and Y can be replaced by an association link (Y correlated to X) or reversed. For a saturated previous block[17], we can create an alternative model by interchanging the relations within this block so that it remains saturated. The other parts of the model remain unchanged.

The following figures show examples of equivalent models used by MacCallum et al. (1993).

Figure 2. Examples of equivalent models.



Initial model

Model 1 (Rules 1 and 2)

Model 2 (Rule 2)

Model 3 (Rule 3)

Model 4 (Rule 3)

X1, X2: Previous block

X3, X4: Focal block

X5: Succeeding block

Source : adapted from MacCallum et al. (1993)

[17] This means that each block variable is connected with all the other variables within the same block.

It is important to point out that these rules do not exhaustively generate all the equivalent models; there are other rules. Breckler (1990), for example, presents a method to generate an equivalent model by introducing reciprocal connections by imposing equality constraints on them.

Although it is possible to generate one or more alternative models, it is important to select models that compete with the initial model. The literature offers criteria for short-listing alternative models worth exploring:

(1) In terms of one or more of the model's theoretical and/or logical foundations (Breckler, 1990), in other words, whether the theoretical or logical interpretation holds. For example, Lee and Hershberger (1990) consider the relationship of a cause to its effect. It is generally acknowledged that causes precede their effects. In this context, all equivalent models that do not comply with this criterion must be discarded; and

(2) By reducing the number of equivalent models through inclusion in the initial model of other variables whose ratios with certain (but not all) variables of the initial model are known by virtue of theory (Stelzl, 1986).

Hershberger (2006) suggest four types of strategies to choose from several equivalent models: (1) the use of indices called *ICOMP* ("*Information complexity criterion*") that make it possible to select models whose estimated parameters are independent of each other, (2) elimination of models with estimation problems, often signs of a model misspecification, (3) selection of models with the highest coefficient of determination *R* (its scope is, on the contrary, limited by the influence of the sample size), and (4) choice of the model with the lowest "*Extended Individual Case Residuals (EICR)*" which is the difference between the reported and predicted values.

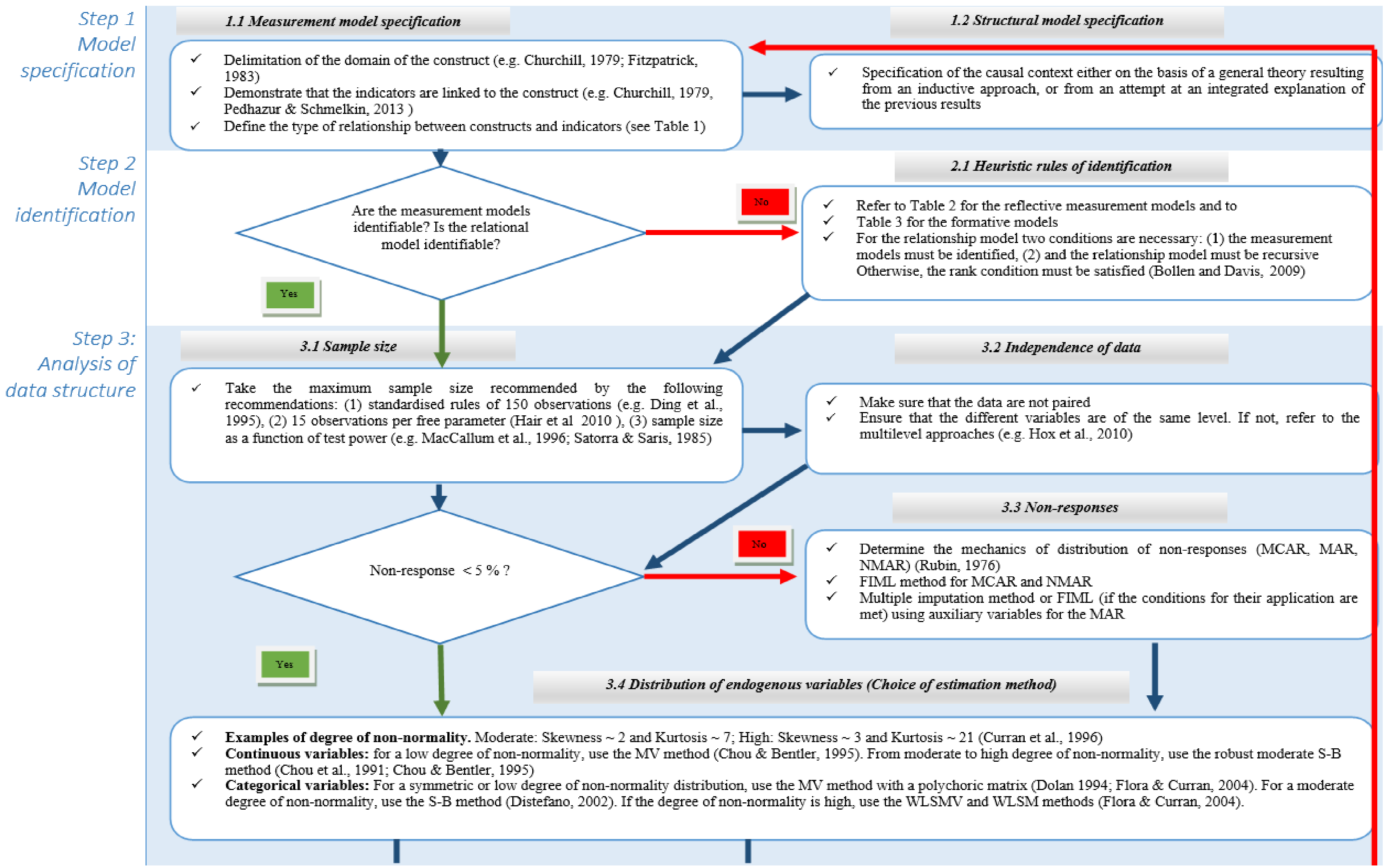
Step 8: Reporting

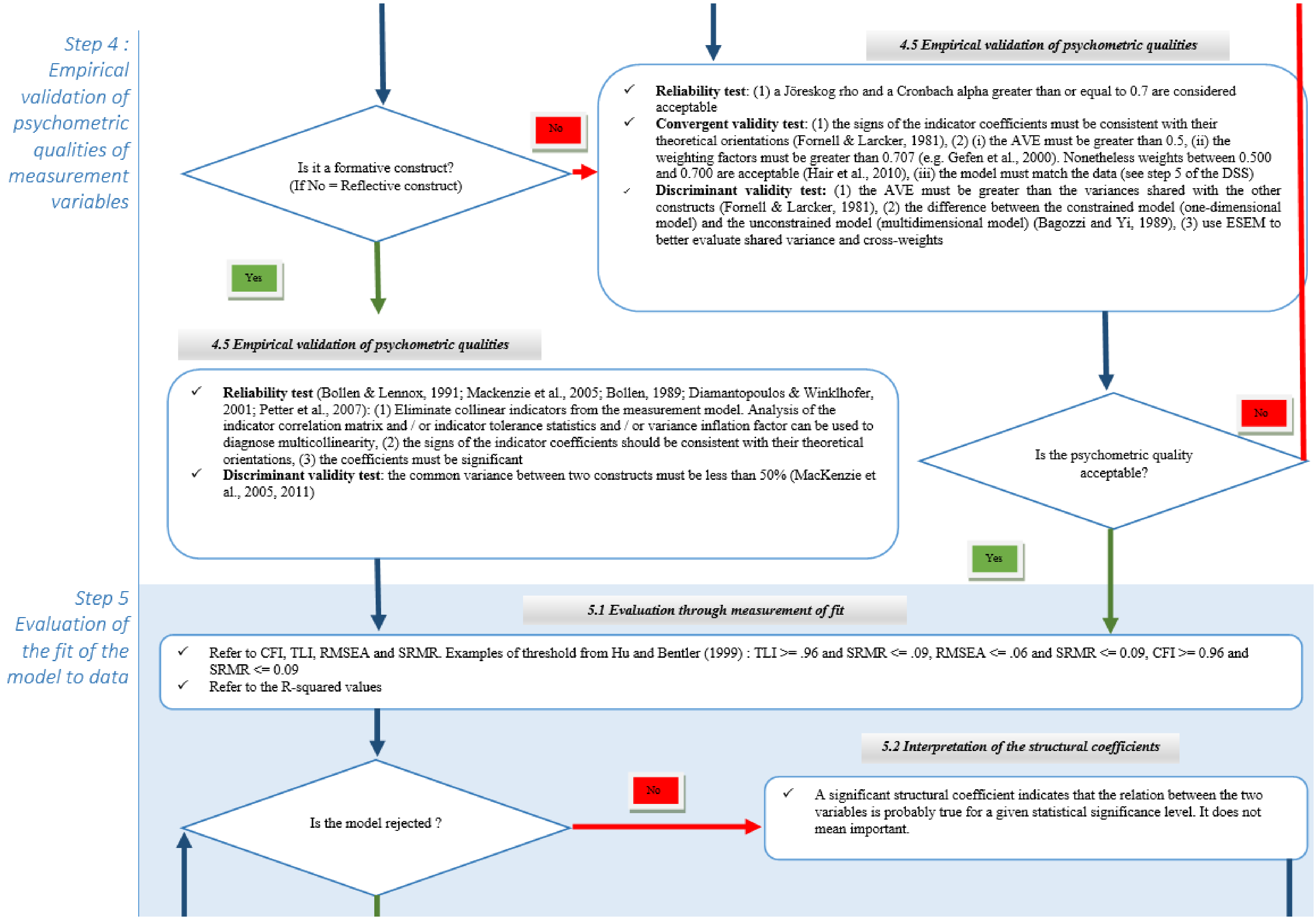
Articles using SEM must provide sufficient details to allow other researchers to replicate their results for external validation (Chin et al., 2008). Chin et al. (2008) considers that the following elements should be particularly reported: the software and the used version, the choice of estimation method, the parameters of the model initially estimated as well as those of the re-specified model (structural coefficients, weights, variances, constraints on the parameters). The initial values chosen for the iterations must be also mentioned.

Finally, we think that is also recommended to postpone the type of the data matrix in order to identify the nature of the study: is it a comparison study (standardized estimates) or a theoretical generalization study?

The following figure (Figure 3) synthesis through a flowchart the various stages and the sub-stages of the decision protocol for the application of SEM.

Figure 3. Decision protocol for research models based on SEM.




Step 4 : Empirical validation of psychometric qualities of measurement variables
Is it a formative construct? (If No = Reflective construct)
No
Yes
4.5 Empirical validation of psychometric qualities
✓ Reliability test: (1) a Jöreskog rho and a Cronbach alpha greater than or equal to 0.7 are considered acceptable
✓ Convergent validity test: (1) the signs of the indicator coefficients must be consistent with their theoretical orientations (Fornell & Larcker, 1981), (2) (i) the AVE must be greater than 0.5, (ii) the weighting factors must be greater than 0.707 (e.g. Gefen et al., 2000). Nonetheless weights between 0.500 and 0.700 are acceptable (Hair et al., 2010), (iii) the model must match the data (see step 5 of the DSS)
✓ Discriminant validity test: (1) the AVE must be greater than the variances shared with the other constructs (Fornell & Larcker, 1981), (2) the difference between the constrained model (one-dimensional model) and the unconstrained model (multidimensional model) (Bagozzi and Yi, 1989), (3) use ESEM to better evaluate shared variance and cross-weights
4.5 Empirical validation of psychometric qualities
✓ Reliability test (Bollen & Lennox, 1991; Mackenzie et al., 2005; Bollen, 1989; Diamantopoulos & Winklhofer, 2001; Petter et al., 2007): (1) Eliminate collinear indicators from the measurement model. Analysis of the indicator correlation matrix and / or indicator tolerance statistics and / or variance inflation factor can be used to diagnose multicollinearity, (2) the signs of the indicator coefficients should be consistent with their theoretical orientations, (3) the coefficients must be significant
✓ Discriminant validity test: the common variance between two constructs must be less than 50% (MacKenzie et al., 2005, 2011)
Is the psychometric quality acceptable?
No
Yes
Step 5 Evaluation of the fit of the model to data
5.1 Evaluation through measurement of fit
✓ Refer to CFI, TLI, RMSEA and SRMR. Examples of threshold from Hu and Bentler (1999) : TLI >= .96 and SRMR <= .09, RMSEA <= .06 and SRMR <= 0.09, CFI >= 0.96 and SRMR <= 0.09
✓ Refer to the R-squared values
5.2 Interpretation of the structural coefficients
Is the model rejected ?
No
✓ A significant structural coefficient indicates that the relation between the two variables is probably true for a given statistical significance level. It does not mean important.

Step 6 Model respecification

6.1 Model respecification

- ✓ Use of the MI and EPC statistics to re-specify the model (see Table 6). However, re-specification should be based solely on theoretical considerations.
- ✓ The results obtained after re-specification (if the model is not rejected) must be replicated on other samples (Cudeck & Brown, 1983). The statistical conclusions regarding the different samples should be similar for the re-specification to be valid.

Step 7 Generalization of the theoretical model

7.1 Replication of results on a new sample

- ✓ Replicate the results obtained on other samples (Steiger, 1990)
- ✓ If it is impossible to use other samples, the researcher may use the cross-validation method. The chi-squared difference method or the cross- validation advocated by Cudeck and Brown (1983) can be performed

7.2 Equivalent model

- ✓ One method is to apply the replacement rules of Lee and Hershberger (1990) and those of Hershberger (2006)
- ✓ To reduce the number of models selected (Breckler, 1990), (1) a priori propose models that have a theoretical foundation and (2) insert other auxiliary variables that reduce the number of equivalent models (Stelzl, 1986)
- ✓ From among the remaining models, select those with independent parameters (ICOMP), which have not encountered estimation problems, have an acceptable R^2, and have the lowest EICR (Hershberger, 2006)

Step 8 Reporting

8.1 Reporting** (Chin et al., *2008)

- ✓ The software and the version used
- ✓ The choice of estimation method
- ✓ The initially estimated model parameters as well as those of the respecified models (structural coefficients, loadings or weighting coefficients), variances and constraints on the parameters). The initial values chosen for the iterations must also be mentioned

2.4 CONCLUSION

This essay seeks to remedy improper use of SEM in the field of consumer behavior research in marketing. To address this issue, it proposed a synthesis of recommendations and experiences, and developed a decision protocol comprised of eight easy-to-understand steps. Each step offers recommendations on SEM best practices in this field of study, and is supplemented by specific articles on SEM from other related disciplines.

Compared to previous article reviews of SEM, which use a tailored case-by-case approach, this decision protocol provides more coherence in applying SEM. This also helps novice researchers avoid the time-consuming need to order the steps themselves. More importantly, each step includes both an explanation of what to do and some directions and illustrations for how to apply the method. The decision protocol provides the broadest thematic variety compared to the other reviews of SEM practices. For example, non-response and prior power analysis were included, a novel step called "generalization of the theoretical model" is incorporated, and more detailed explanations on the problem of variable distributions due to nonnormality (e.g., distinguishing categorical variables from continuous variables) are provided.

The decision protocol can be a useful tool for novice SEM researchers, but can also serve as a reminder for experts or practitioners. Journal editors might for example use this decision protocol to assess the use of SEM in submitted articles. It can also be broadened to apply to other social disciplines using SEM, and can be improved by systematically integrating some novel uses of SEM from other disciplines. In short, the decision protocol given here is open to outside extensions and enhancements. To test the decision protocol in a future study, we plan to do a Delphi survey with a panel of experts and SEM practitioners. This will help us improve the tool's reliability and increase its validity among the SEM community.

2.5 REFERENCES

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49(2), 155-173.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological bulletin*, 103(3), 411- 423.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In M. a. R. E. Schumacker (Ed.), *Advanced structural equation modeling: Issues and techniques* (pp. 243-277). Mahwah, NJ: Lawrence Erlbaum.

Babakus, E., Ferguson Jr, C. E., & Jöreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research*, 24, 222-228.

Bagozzi, R. P. (1977). Structural Equation Models in Experimental Research. *Journal of Marketing Research*, 14(2), 209-226.

Bagozzi, R. P. (2011). Measurement and meaning in information systems and organizational research: Methodological and philosophical foundations. *MIS Quarterly*, 35(2), 261-292.

Bagozzi, R. P., & Fornell, C. (1982). Theoretical concepts, measurements, and meaning. *A second generation of multivariate analysis*, 2(2), 5-23.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the academy of marketing science*, 16(1), 74-94.

Bagozzi, R. P., & Yi, Y. (1989). On the use of structural equation models in experimental designs. *Journal of Marketing Research*, 26(3), 271-284.

Bandalos, D. L., & Finney, S. J. (2010). *Factor analysis: Exploratory and confirmatory*. In The reviewer's guide to quantitative methods in the social sciences (pp. 93-114). New York: Routledge.

Barrett, P. (2007). Structural equation modeling: Adjudging model fit. *Personality and Individual differences*, 42(5), 815-824.

Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International journal of Research in Marketing*, 13(2), 139-161.

Bearden, W. O., Netemeyer, R. G., & Haws, K. L. (2011). *Handbook of marketing scales: multi-item measures for marketing and consumer behavior research*. Thousand Oaks, Calif.: SAGE.

Bernstein, I. H., & Nunnally, J. (1994). *Psychometric theory*. New York: McGraw-Hill.

Biddle, B. J., & Marlin, M. M. (1987). Causality, confirmation, credulity, and structural equation modeling. *Child Development*, 58(1), 4-17.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.

Bollen, K. A., & Davis, W. R. (2009). Causal indicator models: Identification, estimation, and testing. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 498-522.

Bollen, K. A., & Diamantopoulos, A. (2017). In defense of causal-formative indicators: A minority report. *Psychological Methods*, *22*(3), 581-596.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological bulletin*, 110(2), 305-314.

Bollen, K. A., & Pearl, J. (2013). Eight myths about causality and structural equation models. In S. L. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 301–328). New York: Springer.

Box, G., & Cox, D. (1982). An Analysis of Transformations Revisited, Rebutted. *Journal of the American Statistical Association*, 77(377), 209-210.

Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior*, 16(3), 201-213.

Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological bulletin*, 107(2), 260-273.

Brown, R. L. (1989). Using covariance modeling for estimating reliability on scales with ordered polytomous variables. *Educational and Psychological Measurement*, 49(2), 385-398.

Browne, M. W. (1984). Asymptotically distribution - free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37(1), 62-83.

Browne, M. W. (1994). Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods. *Structural Equation Modeling: A Multidisciplinary Journal*, 1(4), 287-316.

Bullock, H. E., Harlow, L. L., & Mulaik, S. A. (1994). Causation issues in structural equation modeling research. *Structural Equation Modeling: A Multidisciplinary Journal*, 1(3), 253-267.

Campbell, D. T. (1960). Recommendations for APA test standards regarding construct, trait, or discriminant validity. *American Psychologist*, 15(8), 546-553.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, 56(2), 81-105.

Chin, W. (1998). Commentary: Issues and Opinion on Structural Equation Modeling. *MIS Quarterly*, 22(1), Vii-Xvi.

Chin, W., Peterson, R. A., & Brown, S. P. (2008). Structural equation modeling in marketing: Some practical reminders. *Journal of marketing theory and practice*, 16(4), 287-298.

Chintagunta, P., Erdem, T., Rossi, P. E., & Wedel, M. (2006). Structural modeling in marketing: Review and assessment. *Marketing Science*, 25(6), 604-616.

Chou, C. P., & Bentler, P. M. (1990). Model modification in covariance structure modeling: A comparison among likelihood ratio, Lagrange multiplier, and Wald tests. *Multivariate Behavioral Research*, 25(1), 115-136.

Chou, C. P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed). *Structural Equation Modeling: Concepts, issues and applications* (pp. 37-55). Thousand Oaks, CA: SAGE.

Chou, C. P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for nonnormal data in covariance structure analysis: a Monte Carlo study. *British Journal of Mathematical and Statistical Psychology*, 44(2), 347-357.

Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16(1), 64-73.

Cohen, J. (1992). A power primer. *Psychological bulletin*, 112(1), 155-159.

Collier, J. E., & Bienstock, C. C. (2009). Model misspecification: contrasting formative and reflective indicators for a model of e-service quality. *Journal of marketing theory and practice*, 17(3), 283-293.

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological methods*, 6(4), 330-351.

Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. Multivariate *Behavioral Research*, 18(2), 147-167.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, 38(4), 529-569.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological methods*, 1(1), 16-29.

Diamantopoulos, A., & Siguaw J. A. (2000). *Introducing Lisrel*. London, UK: SAGE Publications.

Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management*, 17(4), 263-282.

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, 38(2), 269-277.

Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling: A Multidisciplinary Journal*, 2(2), 119-143.

DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(3), 327-346.

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology*, 47(2), 309-326.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(3), 430-457.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 56-83.

Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S., Park, H., & Shao, C. (2016). Applications of structural equation modeling (SEM) in ecological studies: an updated review. *Ecological Processes*, *5*(1), 19.

Finn, A. & Wang, L. (2014). Formative vs. reflective measures: Facets of variation. Journal of Business Research, 67, 2821–2826.

Fitzpatrick, A. R. (1983). The meaning of content validity. *Applied Psychological Measurement*, 7(1), 3-13.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological methods*, 9(4), 466-491.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50.

Garver, M. S., & Mentzer, J. T. (1999). Logistics research methods: employing structural equation modeling to test for construct validity. *Journal of business logistics*, 20(1), 33-57.

Gefen, D., Straub, D., & Boudreau, M.-C. (2000). Structural equation modeling and regression: Guidelines for research practice. Communications of the association for information systems, 4, Article 7.

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research*, 15(2), 186-192.

Gerbing, D. W., & Anderson, J. C. (1992). Monte Carlo evaluations of goodness of fit indices for structural equation models. *Sociological Methods & Research*, 21(2), 132-160.

Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 40–65). Newbury Park, CA: Sage.

Graham, J. W., & Coffman, D. L. (2012). Structural equation modeling with missing data. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 277–295). New York: Guilford Press.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate data analysis (7 ed. Vol. 5): Prentice hall Upper Saddle River, NJ.

Harlow, L. (1985). *Behavior of some elliptical theory estimators with nonnormal data in a covariance structures framework: A Monte Carlo study*. Los Angeles: University of California.

Hershberger, S. L. (2006). The problem of equivalent structural models. In G. Hancock & R. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 13-41). Greenwich, CN: Information Age.

Hox, J. J., Moerbeek, M., & Van De Schoot, R. (2010). *Multilevel ana*lysis: *Techniques and applications*. New York: Routledge.

Hoyle, R. H. (2012). Structural equation modeling with missing data. In R. Hoyle (Ed*.), Handbook of structural equation modeling* (pp. 3–16). New York: Guilford Press.

Hu, L.-T., & Bentler, P. M. (1995). Evaluating model fit. In R. Hoyle (Ed*.*), *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). Thousand Oaks, CA: Sage Publications.

Hu, L. -T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods*, 3(4), 424-453.

Hu, L. -T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.

Hu, L. t., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological bulletin*, 112(2), 351-362.

Hughes, M. A., Price, R. L., & Marrs, D. W. (1986). Linking theory construction and theory testing: Models with multiple indicators of latent variables. *Academy of Management Review*, 11(1), 128-144.

Hulland, J., Chow, Y. H., & Lam, S. (1996). Use of causal models in marketing research: A review. *International journal of Research in Marketing*, 13(2), 181-197.

Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. Sample Size, and Advanced Topics. *Journal of Consumer Psychology*, 20, 90-98.

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of consumer research*, 30(2), 199-218.

Jöreskog, K. G. (1969). Statistical analysis of sets of congeneric tests. *ETS Research Report Series*, 1969(2), i-41.

Jöreskog, K. G., & Sörbom, D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language: Scientific Software International.

Kalleberg, A. L., & Kluegel, J. R. (1975). Analysis of the multitrait-multimethod matrix: Some limitations and an alternative. *Journal of Applied Psychology*, 60(1), 1-9.

Kaplan, D. (1989). Model modification in covariance structure analysis: Application of the expected parameter change statistic. *Multivariate Behavioral Research*, 24(3), 285-305.

Kenny, D. A., Kashy, D., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 233-265). Boston: McGraw-Hill.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(3), 333-351.

Kenny, D. A., & Milan, S. (2012). Identification: A nontechnical discussion of a technical issue. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 145–163). New York: Guilford Press.

Kim, K. H. (2005). The relation among fit indexes, power, and sample size in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 12(3), 368-390.

Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling*. New York: Guilford Press.

Laurent, G., & Kapferer, J.-N. (1985). Measuring consumer involvement profiles. *Journal of Marketing Research*, 22(1), 41-53.

Lee, S., & Hershberger, S. (1990). A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behavioral Research*, 25(3), 313-334.

Lei, P.-W. (2009). Evaluating estimation methods for ordinal data in structural equation modeling. *Quality & Quantity*, 43(3), 495-507.

Lomax, R. G., & Schumacker, R. E. (2012). *A beginner's guide to structural equation modeling*. New York: Routledge Academic.

MacCallum, R. C. (1990). The need for alternative measures of fit in covariance structure modeling. *Multivariate Behavioral Research*, 25(2), 157-162.

MacCallum, R. C. (1995). Model specification: Procedures, strategies, and related issues. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 16-36). Thousand Oaks, CA: Sage Publications.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual review of psychology*, 51(1), 201-226.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological methods*, 1(2), 130-149.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, 114(1), 185-199.

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.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, 35(2), 293-334.

Marcoulides, G. A., & Schumacker, R. E. (1996). *Advanced structural equation modeling: Issues and techniques*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc. Publishers.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological bulletin*, 103(3), 391-410.

Martínez-López, F. J., Gázquez-Abad, J. C., & Sousa, C. M. (2013). Structural equation modeling in marketing and business research: Critical issues and practical recommendations. *European Journal of Marketing*, 47(1/2), 115-152.

Matei, A., & Ranalli, M. G. (2015). Dealing with nonresponse in survey sampling: a latent modeling approach. *Survey Methodology, Statistics Canada*, Catalogue No. 12-001-X, 41(1), 145-164.

McIntosh, C. N. (2007). Rethinking fit assessment in structural equation modeling: A commentary and elaboration on Barrett (2007). *Personality and Individual differences*, 42(5), 859-867.

McQuitty, S. (2004). Statistical power and structural equation models in business research. *Journal of Business Research*, 57(2), 175-183.

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, 22(1-2), 43-65.

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non‐normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38(2), 171-189.

O'cass, A., & Carlson, J. (2012). An empirical assessment of consumers' evaluations of web site service quality: conceptualizing and testing a formative model. *Journal of Services Marketing*, 26(6), 419-434.

O'Leary-Kelly, S. W., & Vokurka, R. J. (1998). The empirical assessment of construct validity. *Journal of operations management*, 16(4), 387-405.

Olsson, U. H., Foss, T., Troye, S. V., & Howell, R. D. (2000). The Performance of ML, GLS, and WLS Estimation in Structural Equation Modeling Under Conditions of Misspecification and Nonnormality. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(4), 557-595.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, Design, and Analysis: An Integrated Approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Peter, J. P. (1981). Construct validity: A review of basic issues and marketing practices. *Journal of Marketing Research*, 18(2), 133-145.

Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, 31(4), 623-656.

Popper, K. (1968). *Conjectures and refutations*. New York: Harper & Row.

Preacher, K. J. (2006). Quantifying parsimony in structural equation modeling. *Multivariate Behavioral Research*, 41(3), 227-259.

Richter, N., Sinkovics, R., Ringle, C., & Schlägel, C. (2016). A critical look at the use of SEM in international business research. *International Marketing Review*, 33(3), 376-404.

Rigdon, E. E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavioral Research*, 30(3), 359-383.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592.

Saris, W. E., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. *Sociological methodology*, 17, 105-129.

Saris, W. E., Satorra, A., & Van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal*, 16(4), 561-582.

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, 50(1), 83-90.

Sempé, L. (2000). Une échelle de mesure de l'appartenance aux cercles sociaux: analyse factorielle confirmatoire multiniveaux. *Recherche et Applications en Marketing* (French Edition), 15(2), 43-57.

Sharma, S., Mukherjee, S., Kumar, A., & Dillon, W. R. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research*, 58(7), 935-943.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25(2), 173-180.

Stelzl, I. (1986). Changing a causal hypothesis without changing the fit: Some rules for generating equivalent path models. *Multivariate Behavioral Research*, 21(3), 309-331.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (Vol. 154, pp. 10-39). Newbury Park, CA: Sage.

Thiétart, R.-A. (2014). *Méthodes de recherche en management*, 4ème édition. Dunod.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 56-75), Sage, Thousand Oaks

Whittaker, T. A. (2012). Using the modification index and standardized expected parameter change for model modification. *The Journal of Experimental Education*, 80(1), 26-44.

Williams, L. J. (2012). Equivalent models: Concepts, problems, alternatives. In R. Hoyle, *Handbook of Structural Equation Modeling* (pp. 247-260). New York: Guilford Press.

Yu, C. (2002). A philosophical investigation of causal interpretation in structural equation models. Paper presented at the Paper presented at the Annual Meeting of the American Education Researcher Association SIG: Structural Equation Modeling, New Orleans, LA.

Zaichkowsky, J. L. (1985). Measuring the involvement construct. *Journal of consumer research*, 12(3), 341-352.

CHAPTER III. ESSAY 2: IMPACT OF SAMPLE SIZE, NONNORMALITY OF VARIABLES AND ESTIMATION METHOD ON FIT INDICES: APPLICATION TO THE CAUSAL MODEL OF CONSUMER INVOLVEMENT

ABSTRACT

Using the Monte Carlo simulation method, this study analyzes the impacts on fit indices by the degree of nonnormality of variables, the sample size, and the choice of estimation method. To address these issues we use the causal model of consumer involvement by Mittal and Lee (1989). Results of this study shows that adjusted goodness-of-fit statistic (AGFI) and Goodness-of-fit statistic (GFI) are not recommended methods to assess the model. Comparative fit index (CFI) and Tucker-Lewis index (TLI) are recommended under specific conditions for the sample size and the choice of estimation methods. Both Root mean square error of approximation (RMSEA) and Standardized root mean square residual (SRMR) combinations do not reject the true model under different settings, and seem to be reliable with maximum likelihood estimation. Finally, the degree of nonnormality seems to have little impact on the fit indices under study.

Keywords: *structural equation methods, fit indices, Monte Carlo simulation, parameter estimation methods, sample size, degree of nonnormality of variables.*

3.1 INTRODUCTION

The concept of involvement has long been a major and controversial topic among consumer behavior researchers in marketing (e.g., Atkinson & Rosenthal, 2014; Bloch, 1983; Mitchell 1979; Zaichkowsky, 1986). Widespread fascination with the concept of involvement is explained by its tight link to major attributes such as brand and product, and to its connection to other concepts, especially its antecedents and consequences (e.g., Laage-Hellman, Lind & Perna, 2014; Bauer, Sauer, & Becker, 2006; Kapferer & Laurent, 1985). Incorporating all of these concepts into the same model allows us to evaluate a product’s consistency, or congruency, with the consumer's needs. Since this topic is highly important for both research and for the practical consequences of that research, the purpose of this study is to replicate the Antecedents – Involvement – Consequences (A-I-C) model of consumer involvement introduced by Mittal and Lee (1989), to show how the choice of fit indices, sample size, degree of nonnormality of variables, and choice of estimation method impact the conclusion of SEM (Structural Equation Modeling) research.

The statistical results based on fit indices, adjusted goodness-of-fit statistic (AGFI), Chi-square, and Goodness-of-fit statistic (GFI) obtained in testing the A-I-C model fail to confirm its validity

(Mittal & Lee, 1989; Flynn & Goldsmith, 1993). The replication that Flynn and Goldsmith (1993) performed of the model confirms the lack of validity of the measurement models used. Nonetheless, such confirmation is by no means universal, since rival studies conclude with non-rejection of the model (e.g., Bezençon & Blili, 2010). An alternative explanation for this inconsistency lies in the manner in which the statistical method is applied to the data analysis. Generally, the A-I-C model was tested by using the Structural Equation Method (SEM), which has the advantage of simultaneously processing all complex relations among all the variables of a model. However, in practice, a model's parameters do not always possess the expected properties (Bandalos, 2006). Quantitative statisticians often encounter situations where the collected data fail to comply with the assumptions underlying the method that they are using.

For instance, in the social sciences, the assumption that the variables are normally distributed is rarely satisfied (e.g., Gierl & Mulvenon, 1995; Harlow, 1985; Micceri, 1989). In the case of SEM, although recommended standards exist for its application, its reliability has not yet been formally determined, in view of the difficulty of finding representative models of social phenomena. For example, researchers usually use maximum likelihood estimation as the standard method for estimating the parameters of structural equations (e.g., Baumgartner & Homburg, 1996; Martínez-López, Gázquez-Abad, & Sousa, 2013). Nonetheless, this estimator's properties are known only asymptotically. As a result, some common rules of thumb — like the recommended sample size of 100 or 150 observations (e.g., Boomsma, 1982), or 5 to 10 observations per estimated parameter (Bentler & Chou, 1987) — may be unsuitable when applying SEM, particularly for models that differ greatly from the original model being tested.

This kind of problem has prompted researchers to propose other manners of applying SEM, such as by means of the Monte Carlo simulation method, in order to obtain more reliable results (e.g., In'nami & Koizumi, 2013, Muthén & Muthén, 2002). Thus, in this study, we apply Monte Carlo simulation to assess the impact of the estimation method, the degree of nonnormality, and the sample size on a range of fit indices in the A-I-C model. The results will help us choose the best fit indices to assess the A-I-C model for future applications.

3.2 LITERATURE REVIEW

Inconsistent Results for the A-I-C Model

Based on various early papers dealing with the concept of consumer involvement (e.g., Bloch & Richins, 1983; Kapferer & Laurent, 1985), Mittal and Lee (1989) proposed a unifying theoretical framework consisting of a causal model of involvement called the A-I-C model of consumer involvement (see Figure 4). The model clearly distinguishes between the different forms, namely the antecedents, and the consequences of a consumer's involvement. In this model, two forms of involvement are considered: product involvement and purchase decision involvement. The former is defined as the interest that the consumer devotes to a category or class of products. Such interest manifests itself in the consumer's perception of the importance of a category of products, or the need to attain a goal, or to conform to certain values. The A-I-C model assumes that low purchasing involvement for a given product category induces random choice of brand. Alternatively, strong involvement entails an expectation of careful brand choice. To link the two forms of involvement, the A-I-C model considers involvement with a product an antecedent to the involvement for the purchasing decision. In the case of an involving product, the purchase decision for the product is not a random one. However, the occurrence of a purchase involvement need not imply that the product is involving.

Mittal and Lee (1989) associate each of these two manners of involvement with four sources of involvement. They were chiefly inspired to do so by the work of Kapferer and Laurent (1985) on the four facets of involvement: (1) perceived importance, (2) perceived risk, (3) symbolic value and (4) pleasure. Understanding the effect of these antecedents allows us to erect a dynamic image of an abstract state and recommend the use of different communication media, depending on which consumer segment is being addressed. More precisely, knowing the influence of these sources on the form of involvement enables us to differentiate and increase the interest shown by a segment of consumers for one class of a particular product. In the A-I-C model, six sources of involvement are retained for the two forms of involvement: (1) Symbolic value, (2) hedonic value, and (3) utility for product involvement; and (1) symbolic value, (2) hedonic value, and (3) risk for brand involvement.

To test the model’s validity, Mittal and Lee (1989) applied it to two types of product categories: clothing (jeans) and electronic devices (video cassettes). For this test, 144 and 136 observations were gathered for jeans and for video cassettes, respectively. The parameters were estimated using the maximum-likelihood method. We highlight two results from Mittal and Lee’s (1989) study concerning the validity of the measurement models and the structural model. For the measurement model, the Chi-squared test rejected the hypothesis of good fit to the data. Nonetheless, the measurement models passed the tests of reliability and validity according to the criteria of Fornell and Larcker (1981). For the structural model, based on the recommended standard thresholds of 0.96 (e.g., Bandalos & Finney, 2010), 0.95 (e.g., Hu & Bentler, 1999), or 0.90 (e.g., Bentler & Bonnet, 1980), the outcome of Mittal and Lee’s (1989) test was rejection of the structural model in both product categories (see Table 8).

Although Mittal and Lee (1989) interpret their results according to the structural parameters, the results for the fit indices may cast serious doubt on their interpretations. Essentially, what is the benefit of interpreting the structural coefficients if the data do not fit the model? According to Olsson, Foss, Troye and Howell (2000), researchers do not know the true values of parameters. Therefore, they need to resort to an empirical fit, utilizing fit indices as a basis for judging the theoretical fit.

A replication of the A-I-C model by Flynn and Goldsmith (1993)—based on a sample comprising 130 observations on “professional clothing”—yields the same conclusion, namely rejection of the model (see Table 8). Here, the authors attribute it to the difficulties that the measurement models encounter regarding convergent and discriminant validity. This finding remains controversial, since subsequent replications of the A-I-C model showed that the models of measurement and structural relation were not rejected statistically (e.g., Bezençon & Blili, 2010). Mittal and Lee (1989) claim that the Chi-squared measure rejects the A-I-C model because of the distribution characteristics of the variables and the effects of sample size.

Accuracy of Fit indices Tests

Several research results show that when facing a high degree of nonnormality, the Chi-squared test is more likely to reject a correct model (e.g., McIntosh, 2007). Other studies conclude that the test often rejects correct models whose sample size is large (Bentler & Bonnet, 1980). Despite the various weaknesses cited in the literature on this fit index, it is almost always published, since it is a historical reference and a basis for calculating most other fit indices (e.g., AGFI, RMSEA, Tucker-Lewis index (TLI), and Comparative fit index (CFI)). Although the finding reached by Mittal and Lee (1989) relies on the Chi-squared test, it should be noted that the other measures yield conflicting results—AGFI rejects the model's goodness of fit to the data (less than 0. 96 or 0.90), unlike SRMR, which is below the recommended threshold of 0.08 (see Table 8).

It was recommended to avoid using the AGFI and GFI measures (Sharma, Mukherjee, Kumar & Dillon, 2005; Kline, 2005). Essentially, these two measures exhibit great variation within each sample-size category (Sharma et al., 2005). This type of result leads researchers to wonder about the choice and characteristics of a good fit index (Boomsma, 2000; Hu & Bentler, 1999; Tanaka, 1993). Breckler (1990) recommends choosing those measures that best evaluate the model's goodness of fit to the data, instead of using only measures that validate the model.

The following fit indices are often recommended in the literature: SRMR, RMSEA, CFI and TLI (e.g., Hu & Bentler, 1999; Lei & Lomax, 2005). Generally, these measures are retained not only because they are relatively insensitive to sample size and to the variables' degree of nonnormality, but also because they are more sensitive than other measures such as the rescaled version of Akaike's information criterion (CAK; Cudeck & Browne, 1983), the cross-validation index (CK; Browne & Cudeck, 1989); the normed fit index (NFI; Bentler & Bonett, 1980), the Bollen fit index (BL86; Bollen, 1986), GFI, and AGFI at detecting mis-specified model (Hu & Bentler, 1998, 1999). For example, Hu and Bentler (1999) proposed using a combination of two indices to better evaluate a model's goodness of fit to the data: TLI $\geq$ 0.96 and SRMR $\leq$ 0.09, RMSEA $\leq$ 0.06 and SRMR $\leq$ 0.09, and CFI $\geq$ 0.95 and SRMR $\leq$ 0.09. Their findings show that these combinations yielded the lowest aggregate rates of type I and type II errors. However, when the sample size is smaller than 250 observations, the CFI-SRMR combination is preferable, since both other combinations tend to reject correct models when the underlying assumptions of nonnormality are violated. This two-index strategy was partially challenged by Fan and Sivo (2005) using a different model than Hu and Bentler (1999). Their results indicated that there is insufficient evidence to support the two-index strategy, and they recommended that researchers should pay attention to overgeneralizing Hu and Bentler's (1998, 1999) findings.

When comparing the various fit indices, the SRMR measure seems less sensitive to violations of the normality assumption and to variations in sample size. It is efficient on condition that the saturation coefficients are large (Anderson & Gerbing, 1984). Another advantage of this fit index results from its standardized nature, which enables comparing its values across different models.

As for the RMSEA, it exhibits the advantages of preferring parsimonious models (Diamantopoulos & Siguaw, 2000), of being relatively less sensitive to poor model specification and to sample size (Fan, Thompson, & Wang, 1999). In addition, it is less influenced by the choice of parameter estimation methods (Fan et al., 1999).

As for the CFI fit index, it too prefers parsimonious models, but in addition has the property of adapting to a model's complexity (Iacobucci, 2010) and is less sensitive to variations in sample size (Byrne, 1998; Fan et al., 1999). On the other hand, it is less effective when there are too many indicators per factor (Kenny & McCoach, 2003).

Like the CFI, the TLI fit index is also less sensitive to variation in sample size (Marsh, Hau, & Wen, 2004). However, it is less efficient at testing a correct model with several variables (Kenny & McCoach, 2003).

The Role of Estimation Methods

Although the choice of fit indices is of great importance, their results depend on which estimation methods are used (Fan et al., 1999; Sugawara & MacCallum, 1993). In the case of variables with a high degree of nonnormality, the maximum-likelihood method, for example, proved to be limited, since the estimators become biased and therefore affect the results for the measures. On these grounds, some authors recommend using alternative estimation methods such as MLM, MLR, and MLMV, which are often deemed robust methods. The MLM method permits correcting the standard errors of the estimated parameters (obtained from the inverse of the asymptotic information matrix) of the maximum-likelihood method in case of a violation in the degree of nonnormality (Curran, West, & Finch, 1996; Satorra & Bentler, 1994). This implies that the parameters are identical in both methods. However, fit indices differ when variables are subjected to high levels of nonnormality, since the adjustment proposed by Satorra and Bentler (1994) consists of adjusting the LR test to the statistic of the mean of a Chi-squared distribution for any given degree of freedom. For the MLMV method, goodness of fit affects not only the mean but also the variance. As a result, the estimated parameters remain identical to those of MV and MLM, but are still different with respect to the fit indices (Maydeu-Olivares & Distefano, 2016). This difference relates only to measures that are calculated by means of the Chi-squared test. For example, the SRMR fit index is not affected. For MLR, another alternative method of obtaining parameters is used, namely the direct maximum-likelihood method. Accordingly, its result differs from the LR test. The statistic used is the T2 * of Yuan, Chan and Bentler (2000). This statistic also follows a Chi-squared law asymptotically; however, it differs from that characterizing the preceding methods. Accordingly, any values thus obtained for fit indices will differ.

Figure 4. The A-I-C model of Mittal and Lee (1989).

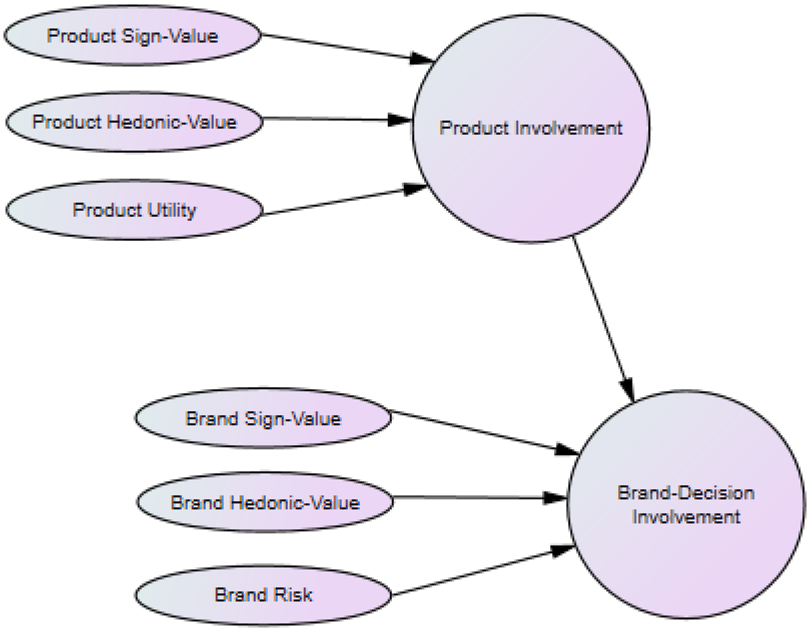


Table 8. Fit indices results for the A-I-C model.

Mittal and Lee (1989) **Products: Videotape**	**Mittal and Lee (1989)** **Products: Clothes**	**Flynn and Goldsmith (1993)** **Products: Clothes**
Chi-squared = 460.17 *df* = 230 *p* = 0.0001 *AGFI* = 0.708 *SRMR* = 0.071	*Chi-squared* = 332.14 *df* = 230 *p* = 0.0001 *AGFI* = 0.785 *SRMR* = 0.063	*Chi-squared* = 612.99 *df* = 230 *p* = 0.0000 *AGFI* = 0.622 *GFI* = 0.711 *SRMR* = 0.259

3.3 METHODS

We used the Monte Carlo simulation method to evaluate the behavior of the A-I-C model based on fit indices. This method should be considered if the problem cannot be solved analytically. Indeed, in SEM, the underlying assumptions are seldom fulfilled (e.g., Gierl & Mulvenon, 1995; Harlow, 1985; Bandalos, 2006). Moreover, most of the sample distribution of the fit indices is unknown (except for the RMSEA and the Chi-test). In such cases, the most suitable analytical tool is Monte Carlo simulation (Bandalos, 2006).

For evaluating the model's goodness of fit to the data, we used the following measures: the Chi-squared test, GFI, AGFI, SRMR, RMSEA, TLI, and CFI. Although the Chi-squared, GFI, and AGFI measures are no longer recommended (Sharma et al., 2005; Kline, 2005), we retained them for the purpose of comparison. Note that they were chosen by Mittal and Lee (1989), and by Flynn and Goldsmith (1993), for evaluating applications of the A-I-C model.

The impacts of the independent variables were analyzed, to wit sample size, the variables' degree of nonnormality, and the method for estimating the parameters on the dependent variable—the fit index. Four sample size categories are considered: 100, 150, 300, and 500 observations. Observations for the first two categories constitute the so-called group of "small sample sizes." On the whole, they are used as rules of thumb for choosing sample size when applying SEM (e.g., Boomsma, 1982). The last two categories are the large-sample groups. Although no rule prescribes the required number of observations for a large sample, and although the concept of a large sample is an asymptotic theory, we arbitrarily defined it as such, since the median sample size in marketing research from 1997 to 2007 is 259 observations (first quartile = 154 observations; third quartile = 419 observations) (Martínez-López et al., 2013), and 178 observations (first quartile = 120 observations; third quartile = 305 observations) from 1977 to 1994 (Baumgartner and Homburg, 1996).

When defining the categories of degree of nonnormality, we followed Curran et al. (1996). These authors suggest the following set of skewness-kurtosis pairs: normal (0, 0), moderately nonnormal (2, 7), and severely nonnormal (3, 21). To generate variables with different degrees of nonnormality, we chose the method of Vale and Maurelli (1983). For the last independent variable, which is choice of parameter estimation methods, we considered the following methods: MV, GLS, MLM (or the corrected Satorra-Bentler estimation method), MLMV, and MLR.[18]

In order to better analyze the results we obtained from the Monte Carlo simulation method, we decided on a three-factor ANOVA. This yields 60 (4 * 3 * 5) independent groups or cells to be analyzed. In our case, we used 1,000 replications to ensure the stability of the results and to limit non-convergence problems. A total of 60,000 replications were made for all groups.

A good method for data generation is to select parameter estimates from prior studies (Paxton, Curran, Bollen, Kirby, & Chen, 2001). Accordingly, we took the estimated parameters from the A-I-C model for clothing (jeans) yielded by the test of Mittal and Lee (1989). The A-I-C model is then considered the true model in this Monte Carlo simulation experiment. Results obtained of the fit indices from the data generated from the simulation can be considered as a threshold criteria to test the model with empirical data. We incorporated interactions between variables into our factorial model, since we wish to evaluate both the impact of the main factors and the interactions among the three other factors. We then calculated the effect size to determine whether the effects of the tested parameters were relatively weak or strong (Cohen, 1988). We only considered results for ANOVA with at least moderate effects—in our case, a variance greater than or equal to 9%.

We then evaluated the power[19] of the A-I-C model fit indices as a function of sample size. The test power was only computed for RMSEA. This fit index not only has the advantage of a known distribution, but also that a conventional method to assess the discrepancy of the null model to the alternative model is provided in the literature (see MacCallum et al., 1996). Although the CFI and Chi-test distribution are also known we did not perform a power test, because to our knowledge no conventional alternative model was proposed in the literature. We also do not specify an alternative model in this study. For the remaining fit indices AGFI, GFI, SRMR, and TLI, power was not computed because their distributions are not known.

We decided that the best computation method would be the one for calculating the test's power as recommended by MacCallum, Brown, and Sugawara (1996), instead of the method of Saris and

[18] We are well aware that other robust estimation methods exist, but we acknowledge that our selection was made arbitrarily. We chose these estimation methods because they are popular in the literature. Although Brown's (1994) ADF is far better known, this method requires at least 1000 observations to be applied. Few studies in the field of consumer behavior research in marketing are capable of gathering such a large sample (Martínez-López et al, 2013).

[19] The probability of rejecting the null hypothesis that the model fits the data, while knowing that the model does not fit the data.

Satorra[20] (1993). Indeed, the selected method is derived from that of Saris and Satorra (1993), but moreover exhibits the advantage of allowing researchers to specify a priori an alternative hypothesis based on the RMSEA, while also letting them determine a priori the required sample size for reaching the desirable power of the test (Kim, 2005).

From among the various tests for hypotheses proposed by MacCallum et al. (1996), we chose the "close fit" method, where the null hypothesis, H0, on RMSEA is different from zero, but sufficiently close to render the model an approximation of the real phenomenon.
Accordingly, we retained the null hypothesis H0: $\varepsilon \leq 0.05$ and the alternative hypothesis H1: $\varepsilon = 0.08$. Hypothesis H1 specifies the threshold at which a model begins to be considered incorrect. For the target power value, we chose the standard value of 0.80. We assigned the following conventional values to alpha: 0.1; 0.05; 0.01 and 0.001.

Lastly, we ran our simulation on R software. Simulated structural equation modeling (SIMSEM) developed by Pornprasertmanit, Miller, Schoeman, Quick, and Jorgensson (2016) enables Monte Carlo simulations of SEM. To analyze variance, we relied mainly on the SPSS software package (except when estimating parameters by the robust maximum-likelihood method, which we did on R software).

3.4 RESULTS

3.4.1 Analysis of aggregate data

Table 9 and Table 10 present the results for the Monte Carlo simulation method. In general, we observed that as sample size grows, the values of the goodness of fit increase. At the same time, for badness of fit, as sample size grows, the values of its measures decline. We also observed that, as the degree of nonnormality of the variables increases, the goodness of fit diminish. For small sample sizes—fewer than or equal to 100 observations—the values of all fit indices, estimated using maximum-likelihood methods, are beneath the recommended threshold of 0.95 (Bandalos & Finney, 2010; Hu & Bentler, 1999). Only CFI and TLI exhibit values above the threshold of 0.90 (Kline, 2005).

[20] The method of Saris and Satorra (1993) requires that the researcher specifies parameters for the alternative model. Performing this task is much more complex.

The results for CFI and TLI, when estimated by the GLS method, comprise values greater than 0.95. We also note fairly specific behaviors for TLI when estimated by the GLS method. Indeed, some results are greater than one, which indicates an overestimation problem. However, CFI generally performs better than TLI. The other two fit indices, AGFI and GFI, fail to attain the recommended threshold of 0.95 (Bandalos & Finney, 2010; Hu & Bentler, 1999), except for GFI in the case of a sample comprising 500 observations. Approximately 300 observations are needed for the AGFI and GFI measures to exceed Kline's (2005) less restrictive lower limit of 0.90. For the two other badness of fit measures—RMSEA and SRMR—we found that RMSEA is usually lower than the recommended threshold of 0.05. Accordingly, this measure performs very well. The same is true of SRMR, for which the estimated values lie beneath 0.08, except in the case of measures for small sample sizes when obtained by the GLS estimation method.

Table 9. Results for measures of goodness of fit via the Monte Carlo simulation method.

Estimation method		***ML***				***GLS***				***MLM***				***MLR***				***MLMV***			
Sample size		**100**	**150**	**300**	**500**	**100**	**150**	**300**	**500**	**100**	**150**	**300**	**500**	**100**	**150**	**300**	**500**	**100**	**150**	**300**	**500**
(*Skewness=0; Kurtosis=0*)	*CFI*	,927	,962	,987	,994	,992	,981	,986	,990	,926	,962	,987	,994	,927	,962	,987	,994	,926	,962	,987	,994
	TLI	,917	,961	,991	,997	1,664	1,141	1,027	1,010	,917	,961	,991	,997	,917	,961	,991	,997	,917	,961	,991	,997
	GFI	,838	,886	,940	,963	,826	,879	,938	,962	,841	,887	,940	,963	,823	,873	,935	,962	,841	,887	,940	,963
	AGFI	,789	,851	,921	,952	,773	,842	,919	,951	,776	,841	,916	,948	,785	,848	,917	,948	,776	,841	,916	,948
(*Skewness=2; Kurtosis=7*)	*CFI*	,920	,957	,984	,992	,988	,979	,980	,987	,920	,957	,984	,992	,920	,957	,984	,992	,921	,957	,984	,992
	TLI	,909	,953	,986	,993	1,509	1,121	1,014	1,001	,909	,953	,986	,994	,909	,953	,985	,993	,909	,954	,985	,994
	GFI	,836	,884	,939	,962	,824	,878	,937	,962	,839	,886	,939	,962	,816	,868	,931	,959	,839	,886	,939	,962
	AGFI	,786	,849	,920	,951	,770	,841	,917	,950	,774	,839	,914	,947	,782	,845	,916	,948	,774	,839	,914	,947
(*Skewness=3; Kurtosis=21*)	*CFI*	,915	,948	,977	,987	,985	,968	,970	,978	,915	,948	,977	,987	,916	,949	,977	,987	,915	,948	,977	,987
	TLI	,903	,943	,976	,987	1,469	1,089	,991	,986	,903	,943	,976	,987	,904	,944	,976	,987	,902	,944	,976	,987
	GFI	,834	,881	,936	,961	,823	,876	,935	,960	,837	,883	,937	,961	,805	,859	,924	,955	,837	,883	,937	,961
	AGFI	,783	,845	,917	,949	,770	,838	,915	,948	,770	,835	,911	,945	,779	,843	,913	,946	,770	,835	,911	,945

Table 10. Results for measures of badness of fit via the Monte Carlo simulation method.

Estimation method		***ML***				***GLS***				***MLM***				***MLR***				***MLMV***			
Sample size		**100**	**150**	**300**	**500**	**100**	**150**	**300**	**500**	**100**	**150**	**300**	**500**	**100**	**150**	**300**	**500**	**100**	**150**	**300**	**500**
(*Skewness=0; Kurtosis=0*)	*RMSEA*	,031	,020	,010	,007	,002	,004	,005	,004	,031	,020	,010	,007	,032	,021	,011	,007	,031	,020	,010	,007
	SRMR	,076	,062	,044	,034	,129	,094	,057	,040	,074	,060	,042	,033	,074	,060	,042	,033	,074	,060	,042	,033
(*Skewness=2; Kurtosis=7*)	*RMSEA*	,034	,023	,012	,009	,003	,005	,006	,006	,034	,023	,012	,008	,036	,024	,013	,009	,033	,023	,012	,008
	SRMR	,077	,063	,044	,034	,130	,095	,057	,041	,074	,060	,043	,033	,074	,060	,043	,033	,074	,060	,043	,033
(*Skewness=3; Kurtosis=21*)	*RMSEA*	,036	,026	,016	,012	,003	,007	,009	,008	,036	,026	,016	,012	,038	,028	,017	,012	,036	,026	,016	,012
	SRMR	,078	,063	,045	,035	,133	,097	,059	,042	,075	,061	,043	,034	,075	,061	,043	,034	,075	,061	,043	,034

3.4.2 Three-way ANOVA interaction analysis

A series of analyses was performed for simple effects and interaction effects on each of the fit indices. Since the experimental design is orthogonal,[21] interpreting the results was a trivial exercise. The tests rejected normality and homoscedasticity (p-value <0.000). Consequently, the robust method of Wilcox (2012) was used to solve the problem of normality of variables. No correction was made for the violation of the assumption of homoscedasticity, since ANOVA is sufficiently robust when the various cells have the same sample size (Glass, Peckham & Sanders, 1972). Table 11 and Table 12 indicate that for each fit indices, the different independent variables are significant, except for TLI with the sample-size variable. The result is quite predictable due to the many observations (60,000 observations for each fit index being analyzed). Even a weak effect of a single explanatory variable could be deemed statistically significant by virtue of its p-value (Levine & Hullett, 2002). Accordingly, we decided to analyze the effect size using eta-squared and partial eta-squared, which are independent of the experimental design (e.g., Lakens, 2013; Levine & Hullett, 2002).

Moreover, the effect size enables deduction of the magnitude of the relationship between two or more variables, which is impossible when using tests of significance. If eta-squared determines the total share of the variance of an explained variable in a model, partial eta-squared by contrast determines the share of variance of the explained variable, while disregarding the variances generated by the other explanatory variables (Levine & Hullett, 2002). This property of partial eta-squared enables comparisons among different models. For purposes of our analysis, if eta-squared and partial eta-squared are both less than 0.09, they are considered negligible even if the Fischer test is significant.

Analysis of simple effects.

Table 11 on eta-squared shows that of the seven fit indices, four are affected by the choice of estimation methods. CFI, AGFI, and GFI are not affected. For the Chi-squared test, the choice of the estimation method is the source of the greatest variance, that is, 15.2%. For the TLI, RMSEA, and SRMR measures, we also found fairly large effect sizes, namely 9.4%, 12.7%, 22.9%, and

[21] All combinations between different categories of factors are represented in the experimental design.

21.3%, respectively. Comparison of partial eta-squared (see Table 12) indicates that SRMR (0.828) is more strongly affected than any other measure, followed by RMSEA (0.18), CHISQ (0.17), and TLI (0.11).

For the effects of the degree of nonnormality variable, the impact remains slight for all measures, compared to the two other independent variables, except for Chi-squared, which is the second most important source of variability. For partial eta-squared, Chi-squared seems to be the one most affected by the degree of nonnormality, while TLI, GFI, and AGFI are the least affected.

The effect of sample size is the principal source of variance for all measures except for Chi-squared and TLI. For all measures, the magnitude is relatively large, namely well above 0.09. The partial eta-squared of the measures reveals that SRMR (0.936), AGFI (0.929), and GFI (0.778) are the measures most influenced by sample size. However, for the CFI, TLI, RMSEA, and SRMR measures, the results for eta-squared and/or partial eta-squared should not be considered definitive, in view of the significance and magnitude of interaction between choice of estimation method and sample size for these measures.

*Analysis of interaction effects***.**

Unlike interactions between choice of estimation method and sample size, the interactions between sample size and the degree of nonnormality, and the interactions between degree of nonnormality and the parameter estimation method, are slight (less than 9% of the variance). It follows that only the results for the interaction between choice of estimation method and sample size deserve to be interpreted for use in the next step.

For the CFI (Figure 5a), we observe wide variation in the fit index within each sample-size category. The interaction effect arises only when the GLS method is used. As a result, the sample-size effect is a major variable for CFIs estimated by means of maximum-likelihood methods. It appears that for a small sample size of 100 observations, GLS yields a CFI better than those obtained from the maximum-likelihood estimation methods. For samples of up to 150 observations, as sample size increases, the CFI drops. Beyond 150 observations, the fit index increases monotonously with sample size.

For TLI (Figure 5b), the interaction is created by the GLS method. The GLS method tends to overestimate (measures greater than one). Failing that, no actual interaction effect appears to exist between sample size and other estimation methods. Therefore, the simple effect—which is predetermined a priori—reflects only the choice of estimation method between either GLS, or else reflects the set of all maximum-likelihood methods combined. Figure 5c shows that robust maximum-likelihood methods are less efficient than conventional methods for small sample sizes with the RMSEA. The measure obtained by means of GLS is the most efficient of all, and the most stable within each sample-size category. It is the switch from a GLS estimation method to a different estimation method that creates the interaction effect. If we exclude the GLS method from the category of estimation methods, we would expect the interaction effect between sample size and choice of estimation method to vanish. If this happened, the choice of sample size would become the primary source of variation in RMSEA. Robust methods present similar performances. Their RMSEAs are all within an acceptable range. The larger the sample size, the smaller the difference among estimation methods.

Analysis of Figure 5d reveals that there is no interaction between parameter estimation methods and sample size (almost parallel curves). Overall, SRMR yields acceptable fit index values for all estimation methods, except for the GLS method in small sample sizes (100–150 observations). Since the interaction effect is absent, the analysis resembles that of aggregated data. Consequently, the size effects of the sample-size variable on SRMR are very significant and are not confused with choice of parameter estimation methods.

Table 11. Results for the eta-squared for the various independent variables (IVs).

Fit indices / IV	CHISQ		CFI		TLI		RMSEA		SRMR		GFI		AGFI	
	F sig.	Eta^2	F sig.	Eta^2	F sig.	Eta^2	F sig.	Eta^2	F sig.	Eta^2	F sig.	Eta^2	F sig.	Eta^2
Estimation Methods (EM)	'***'	0,152	'***'	0,030	'***'	0,094	'***'	0,127	'***'	0,213	'***'	0,008	'***'	0,002
Nonnormality (NN)	'***'	0,028	'***'	0,010	' '	0,002	'***'	0,015	'***'	0,000	'***'	0,001	'***'	0,001
Sample size (SS)	'***'	0,024	'***'	0,230	'***'	0,006	'***'	0,206	'***'	0,647	'***'	0,767	'***'	0,926
EM * NN	'***'	0,001	'***'	0,000	'***'	0,001	'***'	0,001	'***'	0,000	'***'	0,001	'***'	0,000
EM * SS	'***'	0,053	'***'	0,069	'***'	0,138	'***'	0,074	'***'	0,095	'***'	0,003	'***'	0,001
NN * SS	'***'	0,001	'**'	0,000	' '	0,001	'***'	0,000	'***'	0,000	'***'	0,000	'***'	0,000
EM * SS * NN	'***'	0,000	'**'	0,000	'***'	0,002	'***'	0,000	'***'	0,000	'***'	0,000	'***'	0,000

Significance test: 0 '*' 0.001 '**' 0.01 '*' 0.05 '-' 0.1 ' '**

Table 12. Results for the partial eta-squared for the various independent variables (IVs).

FIT INDICES	CHISQ		CFI		TLI		RMSEA		SRMR		GFI		AGFI	
	F sig.	P-eta2	F sig.	P-eta2	F sig.	P-eta2	F sig.	P-eta2	F sig.	P-eta2	F sig.	P-eta2	F sig.	P-eta2
Estimation method (EM)	'***'	0,170	'***'	0,044	'***'	0,110	'***'	0,180	'***'	0,828	'***'	0,037	'***'	0,028
Nonnormality (NN)	'***'	0,036	'***'	0,015	'-'	0,002	'***'	0,026	'***'	0,011	'***'	0,006	'***'	0,010
Sample size (SS)	'***'	0,031	'***'	0,258	'***'	0,008	'***'	0,263	'***'	0,936	'***'	0,778	'***'	0,929
EM * NN	'***'	0,001	'***'	0,000	'***'	0,002	'***'	0,001	'***'	0,003	'***'	0,003	'***'	0,000
EM * SS	'***'	0,067	'***'	0,095	'***'	0,155	'***'	0,114	'***'	0,680	'***'	0,015	'***'	0,014
NN * SS	'***'	0,002	'**'	0,001	' '	0,001	'***'	0,000	'***'	0,001	'***'	0,000	'***'	0,001
EM * SS * NN	'***'	0,000	'**'	0,000	'***'	0,003	'***'	0,000	'***'	0,001	'***'	0,000	'***'	0,000

Significance test: 0 '*' 0.001 '**' 0.01 '*' 0.05 '-' 0.1 ' '**

Figure 5. Interaction effects results for large effect sizes.

Figure 5a.

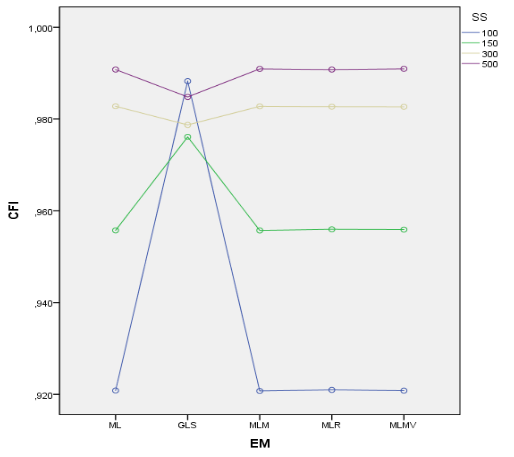


Figure 5b.

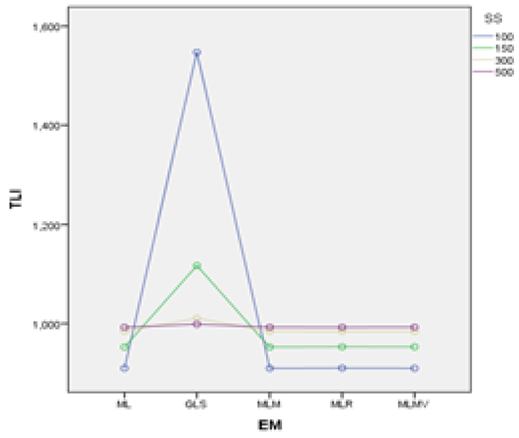


Figure 5c.

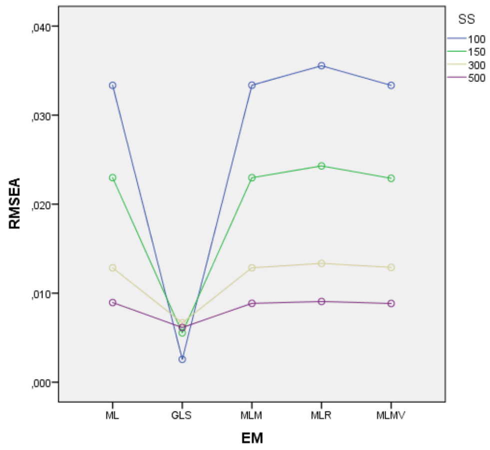
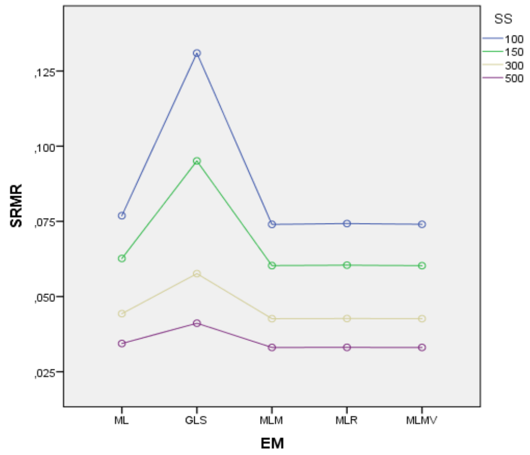


Figure 5d.



3.4.3 Statistical power analysis with RMSEA

We observed from Figure 6 that, if we increase alpha, we need more observations in order to keep the power constant. This relationship between alpha and power is well known. To attain a power of 0.80, Table 13 - variation in power as a function of sample size - shows us that we need 138 additional observations for an alpha of 0.001; 104 additional observations for an alpha of 0.01; 77 additional observations for an alpha of 0.05; and 63 additional observations for an alpha of 0.1. On the other hand, for an alpha of 0.001, we need 138 observations. Accordingly, to test the A-I-C model of Mittal and Lee (1989), a sample size of 143 observations was large enough to attain a power of 0.80. For the smallest sample size in our simulation, that is, 100 observations with an alpha threshold of at least 0.01, the power is already close to 0.80, which is quite high for obtaining a statistically significant result if the model fails to fit the data.

Figure 6. RMSEA's power as a function of sample size.

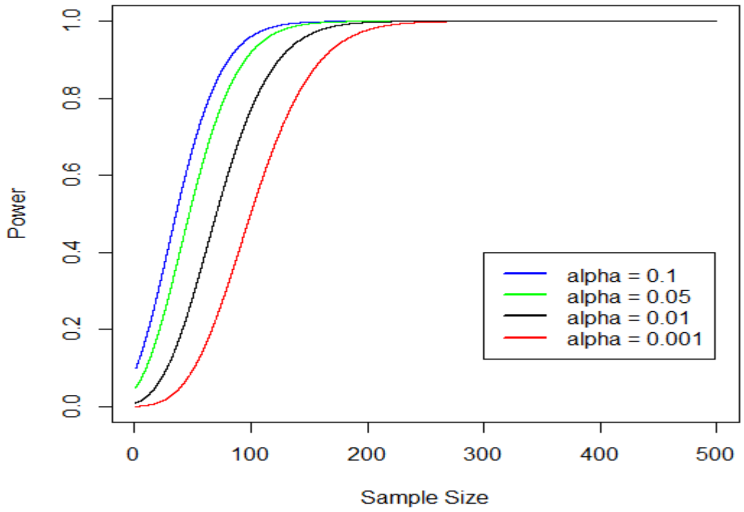


Table 13. Power variation depending on sample size.

Alpha = 0.001	Number of observations (N)	N = 100	N = 138	N = 150	N = 300	N = 500
	Power	0.50	0.80	0.85	0.99	1.00
Alpha = 0.010	Number of observations (N)	N = 100	N = 104	N = 150	N = 300	N = 500
	Power	0.77	0.80	0.96	0.99	1.00
Alpha = 0.050	Number of observations (N)	N = 100	N = 77	N = 150	N = 300	N = 500
	Power	0.91	0.80	0.99	0.99	1.00
Alpha = 0.100	Number of observations (N)	N = 100	N = 63	N = 150	N = 300	N = 500
	Power	0.95	0.80	0.99	0.99	1

3.5 DISCUSSION AND CONCLUSION

Critics of the rules of thumb for fit indices point out that they are influenced by variables such as sample size, estimation methods, and degree of nonnormality. In this study, we reassess the threshold for the A-I-C model developed by Mittal and Lee (1989). Based on this model we generate data from a Monte Carlo simulation, in which the model is considered true. Results of fit indices from this simulation study were considered as an alternative threshold to rule of thumb methods, to assess a further application of this particular model to empirical data. We found that for some indices there is a discrepancy between the rule of thumb criteria and the results from the simulation study.

Specifically, we found that the fit indices, AGFI, and GFI are subject to large variations depending on sample size. Through our simulation, we found that these same fit indices reject the A-I-C model under experimental conditions that are similar[22] to those of Mittal and Lee (1989). As the model generating the data is considered as true under the simulation, these results indicate that it is difficult to attain rule of thumb criteria superior to 0.90, indicating a satisfying model. A sample size of more than 300 observations is needed to reach this threshold for a true model.

The SRMR measure shows good performance in assessing the A-I-C true model in the simulation, even under severe conditions of violation of assumptions (small sample size and / or violation of normality), except with the GLS estimation method. These results corroborate those of prior research, according to which the SRMR is less affected by the degree of violation of the

[22] In terms of sample size category and considering the variables' different degrees of nonnormality.

nonnormality hypothesis and of sample size (Anderson & Gerbing, 1984; Iacobucci, 2010). Nevertheless, this fit index seems to be affected by the choice of methods for estimating parameters. In assessing a replication of A-I-C with empirical data, our simulation results suggest that ML seems to be the best estimation method. It is not affected by sample size or by the degree of nonnormality. For the RMSEA measure, the fit index does not reject the A-I-C model, and it even indicates a very small approximation error. Accordingly, the fit index behaves well when evaluating the model regardless of estimation method and sample size, and induces a result consistent with those determined by Fan et al. (1999). For the CFI, considering the threshold of 0.95, a sample size of less than 150 observations is not recommended to assess the A-I-C model.

Regarding TLI, the GLS estimation method tends to overestimate in some cases, yielding values greater than one. This result is consistent with other research findings, according to which using TLI with GLS leads to overestimation of the fit index (Bentler & Bonnet, 1980; Roussel, Durrieu, & Campoy, 2002), especially with small sample sizes. To assess the A-I-C model, a sample size of no less than 150 observations is needed for TLI to have consistent results. Under the present sample size, the TLI rejects the correct model.

Finally, this study exhibits some limitations. First, we did not have the empirical data or the variance covariance matrix from Mittal and Lee (1989), which would have helped us better replicate their methods and apply different techniques which might affect the results of fit indices. Second, we want to highlight that this study does not aim to evaluate the relevance of the A-I-C model. The relevance of the model is only assessed from empirical data. The results of this study provide prescriptions for the choice of fit indices and the requirements of sample size and estimation method to test A-I-C model replications. Finally, we did not compute power tests for all fit indices selected in this study, as we did not specify alternative models.

3.6 REFERENCES

Anderson, J., & Gerbing, D. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49(2), 155–173.

Atkinson, L., & Rosenthal, S. (2014). Signaling the Green Sell: The Influence of Eco-Label Source, Argument Specificity, and Product Involvement on Consumer Trust. *Journal of Advertising*, 43(1), 33–45.

Bandalos, D. (2006). The Use of Monte Carlo Studies in Structural Equation Modeling Research. In G. Hancock & R. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 385-426). Greenwich, CT: Information Age

Bandalos, D., & Finney, S. (2010). Exploratory and confirmatory factor analysis. In G. Hancock and R. Mueller, (Eds.). *Quantitative Methods in the Social and Behavioral Sciences: A Guide for Researchers and Reviewers* (pp. 93-114). New York, Routledge.

Bauer, H., Sauer, N., & Becker, C. (2006). Investigating the relationship between product involvement and consumer decision-making styles. *Journal of Consumer Behaviour*, 5(4), 342–354.

Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing*, 13(2), 139–161.

Bentler, P., & Bonett, D. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588–606.

Bentler, P., & Chou, C. (1987). Practical Issues in Structural Modeling. *Sociological Methods & Research*, 16(1), 78–117.

Bezençon, V., & Blili, S. (2010). Ethical products and consumer involvement: what's new?. *European Journal of Marketing*, 44(9/10), 1305–1321.

Bloch, P. (1983). Involvement Beyond the Purchase Process: Conceptual Issues and Empirical Investigation. *Advances in Consumer Research*, 9, 413–417.

Bloch, P., & Richins, M. (1983). A Theoretical Model for the Study of Product Importance Perceptions. *Journal of Marketing*, 47(3), 69-81.

Bollen, K. (1986). Sample size and bentler and Bonett's nonnormed fit index. *Psychometrika*, 51(3), 375-377.

Boomsma, A. (1982). Robustness of LISREL against small sample sizes in factor analysis models. In K. Joreskog & H. Wold (Eds.), *Systems under indirection observation: Causality, structure, prediction* (Part I) (1st ed) (pp. 149-173). Amsterdam, Netherlands: North Holland.

Boomsma, A. (2000). Reporting Analyses of Covariance Structures. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(3), 461–483.

Browne, M., & Cudeck, R. (1989). Single Sample Cross-Validation Indices for Covariance Structures. *Multivariate Behavioral Research*, 24(4), 445-455.

Breckler, S. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin*, 107(2), 260–273.

Brown, R. (1994). Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods. *Structural Equation Modeling: A Multidisciplinary Journal*, 1(4), 287-316.

Byrne, B.M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS*. Mahwah, N.J.: Lawrence Erlbaum.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cudeck, R., & Browne, M. (1983). Cross-Validation Of Covariance Structures. *Multivariate Behavioral Research*, 18(2), 147-167.

Curran, P., West, S., & Finch, J. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods,* 1(1), 16–29.

Diamantopoulos, A., & Siguaw, J. (2000). *Introducing Lisrel*. London: SAGE.

Fan, X., & Sivo, S. (2005). Sensitivity of Fit Indexes to Misspecified Structural or Measurement Model Components: Rationale of Two-Index Strategy Revisited. *Structural Equation Modeling: A Multidisciplinary Journal*, 12(3), 343-367.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 56–83.

Flynn, L., & Goldsmith, R. (1993). A causal model of consumer involvement: Replication and critique. *Journal of social behavior and personality*, 8, 129-142.

Fornell, C., & Larcker, D. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39-50.

Glass, G., Peckham, P. and Sanders, J. (1972). Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Review of Educational Research*. 42 (3), 237-288.

Gierl, M., & Mulvenon, S. (1995). Evaluating the application of fit indices to structural equation models in educational research: A review of the literature from 1990 through 1994. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Harlow, L. (1985). *Behavior of some elliptical theory estimators with nonnormal data in a covariance structures framework: A Monte Carlo study*. Los Angeles: University of California.

Hu, L., & Bentler, P. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424-453.

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.

Iacobucci, D. (2010). Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology*, 20(1), 90–98.

In'nami, Y., & Koizumi, R. (2013). Review of Sample Size for Structural Equation Models in Second Language Testing and Learning Research: A Monte Carlo Approach. International *Journal of Testing*, 13(4), 329–353.

Kapferer, J, & Laurent, G. (1985). Consumers' Involvement Profile: New Empirical Results. *Advances in Consumer Research*, 12, 290–295.

Kenny, D., & McCoach, D. (2003). Effect of the Number of Variables on Fit indices in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(3), 333–351.

Kim, K. (2005). The Relation Among Fit Indexes, Power, and Sample Size in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 12(3), 368-390.

Kline, R. (2005). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Laage-Hellman, J., Lind, F. & Perna, A. (2014) Customer Involvement in Product Development: An Industrial Network Perspective. *Journal of Business-to-Business Marketing*, 21(4), 257-276

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 1–12

Lei, M. & Lomax, R. (2005). The Effect of Varying Degrees of Nonnormality in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 12(1), 1- 27.

Levine, T., & Hullett, C. (2002). Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research. *Human Communication Research*, 28(4), 612–625.

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130–149.

Marsh, H., Hau, K., & Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(3), 320–341.

Martínez-López, F., Gázquez-Abad, J., & Sousa, C. (2013). Structural equation modeling in marketing and business research. *European Journal of Marketing*, 47(1/2), 115–152.

Maydeu-Olivares, A., & DiStefano, C. (2016). Maximum likelihood estimation of structural equation models for continuous data: Standard errors and goodness of fit. *Working paper*.

McIntosh, C. (2007). Rethinking fit assessment in structural equation modeling: A commentary and elaboration on Barrett (2007). *Personality and Individual Differences*, 42(5), 859–867.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105(1), 156–166.

Mitchell, A. (1979). Involvement: A Potentially Important Mediator of Consumer Behavior. *Advances in Consumer Research*, 6, 191–196.

Mittal, B., & Lee, M. (1989). A causal model of consumer involvement. *Journal of Economic Psychology*, 10(3), 363–389.

Muthén, L., & Muthén, B. (2002). How to Use a Monte Carlo Study to Decide on Sample Size and Determine Power. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(4), 599-620.

Olsson, U. H., Foss, T., Troye, S. V., & Howell, R. D. (2000). The Performance of ML, GLS, and WLS Estimation in Structural Equation Modeling Under Conditions of Misspecification and Nonnormality. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(4), 557-595

Paxton, P., Curran, P., Bollen, K., Kirby, J., & Chen, F. (2001). Monte Carlo Experiments: Design and Implementation. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(2), 287–312.

Pornprasertmanit, S., Miller, P., Schoemann, A., Quick, C., & Jorgensen, T. (2016). SIMulated Structural Equation Modeling. Package 'simsem' on R software. Retrieved August 25, 2016, from https://cran.r-project.org/web/packages/simsem/simsem.pdf

Roussel, P., Durrieu, F., & Campoy, E. (2002). *Méthodes d'équations structurelles: recherches et applications en gestion*. Paris: Economica.

Saris, W., & Satorra, A. (1993). Power evaluations in structural equation models. In K. Bollen and J. Long (Eds.), *Testing structural equation models* (pp. 181-204). Newbury Park, CA: Sage.

Satorra, A., & Bentler, P. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. Von Eye and C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Sharma, S., Mukherjee S., Kumar, A. and Dillon, W. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research*, 58(7), 935-943.

Sugawara, H., & MacCallum, R. (1993). Effect of Estimation Method on Incremental Fit Indexes for Covariance Structure Models. *Applied Psychological Measurement*, 17(4), 365–377.

Tanaka, (1993). Multifaceted conceptions of fit in structural equation models. In K.A. Bollen and J.S. Long, (Eds.), *Testing structural equation models* (pp. 10-39). Newbury park: CA: Sage Newbury park, CA: SAGE.

Vale, C., & Maurelli, V. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, 48(3), 465–471.

Wilcox, R. (2012). *Introduction to Robust Estimation and Hypothesis Testing* (3rd ed). New York, NY: Academic Press, Elsevier.

Yuan K. H., Chan W., Bentler P. M. (2000). Robust transformation with applications to structural equation modeling. *The British Journal of Mathematical and Statistical Psychology*. 53, 31–50.

Zaichkowsky, J. (1986). Conceptualizing Involvement. *Journal of Advertising*, 15(2), 4–34.

3.7 APPENDICES

Appendix 1. Sample application of the "sim" function.

A description of the "sim" function for applying the Monte Carlo method can be found in the R documentation. Using the "sim" function requires prior installation of the "simsem" package. The following R codes show an example of the application of the "sim" function with structural equations derived from the A-I-C model.

```
require(simsem)

# Population model
popModel                    <-   'f2 =~ 0.78*y1
                                  f2 =~ 0.86*y2
                                  f2 =~ -0.74*y3
                                  f1 =~ 0.74*y4
```

```
f1 =~ 0.85*y5
f1 =~ 0.64*y6
f3 =~ 0.82*y7
f3 =~ 0.86*y8
f3 =~ 0.84*y9
f4=~ 0.90*y10
f4=~ 0.87*y11
f4=~ 0.78*y12
f5=~ 0.68*y13
f5=~ 0.78*y14
f5=~ 0.71*y15
f6=~ 0.84*y16
f6=~ -0.73*y17
f6=~ 0.87*y18
f7=~ 0.77*y19
f7=~ 0.75*y20
f7=~ -0.74*y21
f8=~ 0.87*y22
f8=~ -0.60*y23
f8=~ 0.89*y24

f1 ~ 0.24*f2 + (-0.10)*f6 + 0.49*f7 + 0.295*f8
f2 ~ 0.16*f3 + 0.14*f4 + 0.72*f5
f2 ~~ 0.16*f2
f1 ~~ 0.32*f1        '
# Analysis model
analysisModel        <-    'f2 =~ y1
```

```
f2 =~ y2

f2 =~ y3

f1 =~ y4

f1 =~ y5

f1 =~ y6

f3 =~ y7

f3 =~ y8

f3 =~ y9

f4=~ y10

f4=~ y11

f4=~ y12

f5=~ y13

f5=~ y14

f5=~ y15

f6=~ y16

f6=~ y17

f6=~ y18

f7=~ y19

f7=~ y20

f7=~ y21

f8=~ y22

f8=~ y23

f8=~ y24


f1 ~ f2 + f6 + f7 + f8

f2 ~ f3 + f4 + f5

f2 ~~ 0.16*f2
```

```
f1 ~~ 0.32*f1 '

Output <- sim(1000, model=analysisModel, n=150, generate=list(model = popModel, skewness = rep(3,24), kurtosis = rep(21,24)), std.lv=TRUE, lavaanfun = "sem",estimator = "wlsm")

summary(Output)

design4_3_2 <- data.frame(CHISQ = Output@fit$chisq, CFI = Output@fit$cfi, TLI = Output@fit$tli, RMSEA = Output@fit$rmsea, SRMR = Output@fit$srmr, GFI = Output@fit$gfi, AGFI = Output@fit$agfi)
```

Appendix 2. Example of calculating the power of the RMSEA test on R.

Illustration for:

- a "close fit" test: h0 = 0.05 against h1 = 0.08,
- a degree of freedom of 230,
- an alpha of 0.1
- sample size of 200

```
#Alpha level : alpha

alpha = 0.1

#Degree of freedom : df

df = 230

#Sample size: n

n = 200

#Close fit test

#Null hypothesis: h0

#Alternative hypothesis: h1

h0 = 0.05

h1 = 0.08
```

```
#Power calculation
#Non-centrality parameter: ncp
#ncp0: ncp under the null hypothesis
#ncp1: ncp under the alternative hypothesis
ncp0 = (n-1)*df*(h0)^2
ncp1 = (n-1)*df*(h1)^2
# Example of close fit test
#Quantiles : q
q = qchisq(alpha,df,ncp=ncp0,lower.tail=F)
rmseaPower = pchisq(q,df,ncp=ncp1,lower.tail=F,log.p=F)
(rmseaPower)
```

Appendix 3. Results of the Kolmogorov-Smirnov normality test.

Indices	Statistics	df	p-value
CFI	0,311	60000	0,000
TLI	0,213	60000	0,000
RMSEA	0,355	60000	0,000
SRMR	0,114	60000	0,000
GFI	0,124	60000	0,000
AGFI	0,120	60000	0,000

Appendix 4. Example of applying Wilcox (2012) robust method on R.

R codes for AGFI:

```
Require(WRS2)

> t3way(formula = as.numeric(AGFI) ~ as.factor(EM) * as.factor(NN) * as.factor(SS), data = fullData2)
```

Call:

t3way(formula = as.numeric(AGFI) ~ as.factor(EM) * as.factor(NN) * as.factor(SS), data = fullData2)

	Value	p-value
as.factor(EM)	3610.802	1e-04
as.factor(NN)	108674.080	1e-04
as.factor(SS)	912898.042	1e-03
as.factor(EM):as.factor(NN)	128081.918	1e-03
as.factor(EM):as.factor(SS)	8297.563	1e-03
as.factor(NN):as.factor(SS)	10641.188	1e-03
as.factor(EM):as.factor(NN):as.factor(SS)	12641.994	1e-03

CHAPTER IV. ESSAY 3: CONFIRMATORY FACTOR ANALYSIS AND EXPLORATORY STRUCTURAL EQUATION MODELING: THE CASE OF AN ASSESSMENT OF FOUR MEASUREMENT MODELS OF CONSUMER INVOLVEMENT

ABSTRACT

The aim of this paper is to compare the relative performance of two statistical structural equation method (SEM) tools used for operationalizing constructs: confirmatory factorial analysis (CFA) and exploratory structural equation modeling (ESEM). This comparison was empirically performed on four measurement models for consumer involvement. The four selected measurement models are the personal involvement inventory for advertising (PIIA), the consumer involvement product (CIP), the revised-revised personal involvement inventory (RRPII), and the modified personal involvement inventory (modified PII). Criteria for the comparison are based on psychometric quality, ease of use, and the models' fit to the data. Results showed divergence in regards to discriminant validity between the two methods for the RRPII and CPI tools. It is also showed that ESEM is not a substitute for CFA, but should be instead considered as its complement. Finally, as exploratory structural equation modeling is a recent approach within SEM, to the best of our knowledge this approach has not yet been applied in marketing studies of consumer behavior.

Keywords: *structural equation methods, confirmatory factor analysis, exploratory structural equation modeling, measurement models, consumer involvement.*

4.1 INTRODUCTION

From an historical standpoint, from 1960 to 2011, the only methods for validating measurement models were those of exploratory factor analysis (EFA) and confirmatory factorial analysis (CFA). EFA is an approach to developing measurement models, used to determine the number of latent factors underlying the manifest/observed variables to condense data utilizing retained remaining factors and to identify their significance theoretically, either as already acquired or to be developed. Its limits are imposed by its attributes, which include a predetermined guided approach and randomly setting the number of latent factors. Mulaik (1987) states that EFA is a method for developing hypotheses, while CFA follows as a logical consequence of the former to confirm the hypotheses. As such, CFA imposes a priori constraints on the theoretical relations among variables (Bollen & Pearl, 2013). The discrepancy between the postulated model and data thus enables us to decide whether to reject the model or not. More recently, correlation coefficient between observed and latent variables considered equal to zero would lead to an overly restrictive hypothesis[23] (Marsh, Liem, Martin, Morin & Nagengast, 2011). In most cases, the rejection of a theoretical model by the data forces the researcher to resort to the method of ‘modification indices’, which avoids rejection of the null hypothesis. The approach then becomes exploratory rather than confirmatory (Bollen, 1989; Browne, 2001). However, the use of this method is the subject of several controversies on model validity, as it is likely that using this methodology to adjust the model is a matter of luck (Brannick, 1995; McCallum & Austin, 2000).

Additionally, other studies have shown that rigorously imposing zero values in this manner induces over-estimation for inter-factor correlations. For example, on evaluating the measures of sports coaches' interpersonal styles, Stenling, Ivarsson, Hassmén and Lindwall (2015) compare the results of CFA and ESEM. CFA yields inter-factor correlations between 0.956 and 0.992, while those of ESEM are between 0.473 and 0.708. This raises doubts about the validity of constructs and might also be a source of multicollinearity in models where these constructs are treated as explanatory variables (Marsh et al., 2011, 2010; Cole, Ciesla & Steiger, 2007). This may result in biased structural parameters, which would then alter the predictive power of the constructs (Marsh et al., 2011, 2010; Asparouhov & Muthén, 2009; Cole et al., 2007).

Nonetheless, ESEM has been applied mainly in psychology research for evaluating multi-dimensional models in higher education (Marsh et al., 2009) and personality traits (Marsh et al., 2010; Rosellini & Brown, 2010). This method allows us to better investigate concepts that are theoretically approximated due to its ability to enable the estimation of crossed items in a multidimensional measurement model.

To the best of our knowledge, ESEM has never been applied to the study of consumer behavior research in Marketing. To evaluate the performance of this tool, we chose to replicate the four measurement models of consumer involvement: personal involvement inventory for advertising (PIIA), revised - revised personal involvement inventory (RRPII), consumer involvement profiles (CIP), and Mittal’s (1995) modified PII. We selected this concept because of its importance in explaining consumer choice for one product comparatively Ever since the involvement of the concept was propagated by Krugman (1965), its importance has elicited a significant amount of research. Despite the interest it awakens, involvement is also extremely controversial, explaining

[23] The CFA model, on the other hand, supports the model’s parsimony.

why it presently has no generally accepted definition of the concept (e.g. Lastovicka & Gardner, 1979; Zaichkowsky, 1986; Houston & Rothschild, 1978). The plethora of existing definitions of 'involvement' might thus provide an enriched view of the involvement construct. However, in practice, the more definitions there are of concepts or constructs, the more closely they must match measurement models to be able to manipulate them (Rossiter, 2002). Accordingly, researchers must precisely circumscribe what is contained within the definition and exclude what is not (Churchill, 1979). As such, it is not unusual to find cases where the domain of the construct is not well defined or remains unclear (Andrews, Durvasula, & Akhter., 1990). Such a situation can be caused by the fact that the same item was selected to reflect different constructs emerging from closely related concepts. Moreover, impacts can be detected in the dimensional structure, and the domain of the consumer involvement construct is affected by this problem (e.g. McQuarrie & Munson, 1992; Jain & Srinivasan, 1990; Mittal, 1995). On the same definition basis, some authors perceive involvement as one-dimensional (e.g. Zaichkowsky, 1985), and others as multidimensional (e.g., McQuarrie and Munson, 1992). In this context, we replicate across product categories the measurement models derived from the work of Laurent and Kapferer (1985), Zaichkowsky (1994), McQuarrie and Munson (1992), and Mittal (1995). By means of this replication, we compare the performance of CFA with that of ESEM.

4.2 LITERATURE REVIEW

4.2.1 Exploratory structural equation modeling (ESEM)

The recent introduction of ESEM alleviated the inherent limitations of EFA and CFA (Marsh et al., 2009). The ESEM incorporates both EFA properties (e.g. oblique and orthogonal rotation of factor axes, estimates of standard errors) and CFA properties (e.g. fit indices, factor regression). However, unlike CFA, ESEM enables estimating without restriction all the factor contributions of a measurement model.

One of ESEM's major breakthroughs was to allow for the standard errors of coefficients using the method proposed by Jennrich (2007): it allows us to test the significance of each estimated coefficient and also to evaluate inter-factor links if they are orthogonal or oblique. Accordingly, this test enables replacing arbitrary decision rules to eliminate the small weights required by the EFA and avoid any a priori setting of zero values for certain relations as formulated in CFA.

According to Hu and Bentler (1998), when one or more parameters are constrained to equal zero whereas the real values are not zero, the model is poorly specified and under-parameterized. This problematic case can arise especially for multidimensional measurement models (Perry et al., 2015).

Moreover, although on one hand the EFA permits identifying the factor structures of a construct, it is unable to perform an invariance test by means of different groups. Moreover, if CFA allows it, it overestimates both inter-factor correlations and fit indices. Consequently, its estimates of these magnitudes are unacceptable for factor structures well defined by EFA. This is a direct consequence of the overly strong hypothesis applied to the CFA, according to which items should be located on a single factor (Marsh et al., 2011). As a result, by combining EFA and CFA, ESEM can solve the problem. On the basis of research on Monte Carlo simulations, Asparouhov and Muthén (2009) show that the poor specification of measurement models might have a significant impact on the rest of the model.

In particular, Asparouhov and Muthén (2009) show that eliminating small weights in the CFA can generate structural parameter bias. However, this also makes the fit indices chi-square, RMSEA, CFI, and TLI unreliable. Their examples of simulating 500 samples of size 1000 show that the chi-square for the CFA 100% rejects the true data generated from simulation of the true correct SEM model, compared to a 7% rejection rate for the ESEM. They conclude that the omission of small weights in the CFA can lead to a rejection of a true correct SEM model. Only with an SRMR fit index for values of less than 0.08 is the CFA model not rejected (Asparouhov & Muthén, 2009).

Nonetheless, one of the limitations of the ESEM arises from the problem of the factor rotation procedure, which causes indeterminacy of factor contributions and inter-factor correlations (Sass and Schmitt, 2010). These two parameters vary depending on which rotation procedure is selected (Sass and Schmitt, 2010). Therefore, several equivalent models that have the same goodness-of-fit might be obtained (Schmitt & Sass, 2011). Furthermore, there is hitherto no consensus within the scientific community on the optimal choice of rotation criteria. Until such a consensus emerges, it is recommended to explore several types of factor rotation methods (e.g. Asparouhov and Muthén, 2009; Sass and Schmitt, 2010; Jennrich, 2007).

Another limitation of ESEM is that it is seldom applied because it is very recent. No study has yet verified how relevant the thresholds of the fit indices are (Marsh et al., 2010, 2009). Therefore, Marsh et al. (2010, 2009) suggest that measures of parsimony such as TLI and RMSEA would be more responsive to ESEM's characteristics, because numerous parameters would be estimated.

4.2.2 Models for measuring involvement: CIP, PIIA, RRPII and modified PII

The generic measures of the 'consumer involvement' construct principally involve two models that are well known in the literature: Laurent and Kapferer's (1985) on CIP and Zaichkowsky's (1985) (PII).

Starting from these two measurement models, other equally important ones were derived, such as the RRPII, developed by McQuarrie and Munson (1992), and the modified PII, developed by Mittal (1995). Although the aims of Laurent and Kapferer (1985) and Zaichkowsky (1985) were to allow generic measurement of consumer involvement, the dimensional structures of the two measurements differ.

As such, Zaichkowsky (1985) proposes a one-dimensional model of involvement, whereas Laurent and Kapferer (1985) propose a model with five facets (or dimensions) of involvement. The facets of the CIP are the following: (i) personal interest in the product category, (ii) perceived importance of the negative consequences of a wrong choice, (iii) subjective probability of making the wrong choice, (iv) sign value, and (v) hedonic value assigned by consumer to the product. Zaichkowsky (1985) defines involvement as the way in which someone perceives an object as personally relevant to their needs, values, and interests. She develops a generic measurement model adapted to attributes such as product, advertising, and purchase intention. A 20-item measurement model is thus generated for assessing consumer involvement in all contexts. Consequently, the practical side of the model, due to its one-dimensional structure, facilitates comparing product categories. Moreover, this model has become an essential resource within the scientific community despite the fact it is criticized for its one-dimensional structure (e.g. McQuarrie and Munson, 1992; Mittal 1995).

Moreover, in response to these criticisms, Zaichkowsky (1994) revises her work, proposing more parsimonious – and consequently simpler – measurement models, particularly the PIIA. The number of items is reduced to 10 by eliminating redundancy. Conversely, she developed these consumer involvement models specifically to study involvement in advertising; but, they are nevertheless generic. Zaichkowsky (1994) identifies two dimensions she calls 'emotional relevance' and 'rational relevance'. Each dimension comprises five items: the affective dimension comprises interest, attraction, fascination, passion, and involvement; the cognitive dimension comprises importance, relevance, meaning, value, and need. Although the two-dimensional structure is identified, additional replications are still needed to compare the two-dimensional and one-dimensional structures.

To improve PII, Mittal (1995), relying on a broad consensus in the literature, redefines consumer involvement by suggesting that involvement can partly reflect product importance. He then defines involvement as the way in which somebody perceives an object as personally important in terms of his or her own needs, values, and interests. Therefore, he states that the PII has a multidimensional structure comprising four factors: (i) importance, (ii) relevance, (iii) hedonism, and (iv) attitude. Moreover, Mittal (1995) considers the importance factor to be the only manifestation of involvement, and proposes a modified version of PII. As such, unlike Zaichkowsky (1986, 1994), Mittal (1995) distinguishes the concept of importance from that of relevance; consequently, an object is relevant when it fulfils a function (needs, values etc.), but fulfilling a function does not necessarily render the object important. For Mittal (1995), the hedonism factor is not part of involvement, because an object can possess hedonic character for someone, without necessarily being important. In some cases, hedonic value can even be an antecedent to involvement. Mittal (1995) conceptually distinguishes the attitude factor from the involvement factor, describing attitude as an attraction/avoidance predisposition to an object or activity. Involvement, by contrast, is a mindset of indifference/concern.

In the literature, other authors, namely McQuarrie and Munson (1992), introduce the RRPII, a modified PII reduced to only 10 items. These authors claim that the PII constructs have low validity, but also that the PII is difficult to apply because of its length. They found that several items reflected concepts other than involvement, namely attitude and pleasure. Accordingly, the model is not one-dimensional. Moreover, in the line of Kapferer and Laurent (1985), they claim that the PII is incapable of reflecting the different types of involvement described in literature. Then, they expanded the conceptualization underlying the PII by considering two main latent factors of involvement: the 'cognitive' and the 'affective' factor. They use the terms 'importance' and 'interest' to respectively designate these two manifestations of involvement. In contrast to most researchers (e.g. Mitchell, 1979; Rothschild, 1984; Zaichkowsky, 1985), who emphasize either one of these factors or the other, McQuarrie and Munson (1992) insist that involvement cannot be confined solely one dimension. Consequently, the involvement measurement items must be sampled on the basis of these two factors. Their findings revealed that the 'interest' facet of the involvement construct is incapable of predicting the range of behaviours associated with involvement in the case of headache medicines. Similarly, the 'relevance' facet has low explanatory power for involvement in categories of products such as jeans.

Therefore, PII is not the only instrument criticized for its dimensional structure. For example, criticism of the CIP, the other major instrument, focuses on the difference between the antecedents of involvement and involvement itself. According to Mittal (1989, p. 697):

> "*If we assume those four factors to be antecedents, then we have measured only antecedents and not involvement itself. On the other hand, if we are to consider the 4 factors as facets, then a multi-faceted view of involvement is implied"*.

Thus, this study concludes that only the 'importance' dimension can be deemed representative of involvement. Consequently, the other dimensions are classified in the antecedent category and cannot represent involvement.

Table 14 below shows the CIP, PIIA, RRPII, and modified PII measurement models. The dimensions that are predefined by each model are highlighted, as are the items deleted from or added to the PII in the PIIA, RRPII, and modified PII models.

On one hand, comparing these four measurement models would allow determining their relative performance in terms of psychometric qualities, such as reliability and validity. On the other hand, replicating these different measurement models enables confirming or invalidating their dimensional structure across other product categories. Therefore, the following hypotheses were adopted:

- On the CIP model:

H0a: The CIP model consists of five dimensions: (i) interest, (ii) perceived importance (iii) subjective probability of making the wrong choice, (iv) sign value, and (v) hedonic value.

Against

H1a: The CIP model is not composed of five dimensions.

- On the PIIA scale:

H0b: The PIIA model is composed of two dimensions: 'emotional relevance' and 'rational relevance'.

Against

H1b: The PIIA model is one-dimensional.

- On the modified PII model:

H0c: The modified PII model is one-dimensional

Against

H1c: The modified PII model is not one-dimensional.

- On the RRPII model:

H0d: The RRPII model is composed of two dimensions: 'importance' and 'interest'.

Against

H1d: The RRPII model is one-dimensional.

4.3 METHODOLOGY

4.3.1 Data collection

The measurement models and hypotheses being tested were validated empirically with data on footwear products. The English questionnaire from the PIIA, modified PII, and RRPII models was translated into French by two translation experts at the University of Neuchâtel: Dr. Suzana Zink of the Language Centre and Dr. Katarzyna Jagodzinska of the Enterprise Institute. Convenience samples are used to compare the four measurement models.

The surveyed population was the student body of the University of Neuchâtel (Switzerland). The university has a total of 4376 students, and data collection was performed in several classrooms during the spring semester of 2016. The teachers responsible for each class granted us permission to conduct the survey. Information on the classrooms and the campus are given in the Appendix 5. The questionnaires were administered on paper, and respondents were given 15 minutes to answer the questions.

Overall, 15.35% of students at this university study in the Faculty of Economics and Business, 18.53% in the Faculty of Law, 17.26% in the Faculty of Science, and 48.24% in the Faculty of Literature and Human Sciences. 512 usable filled-out questionnaires were collected from the respondents, representing 11.70% of the target population. The distribution among the different faculties was as follows: 14.06% for the Faculty of Economics and Business, 25.78% for the Faculty of Law, 30.66% for the Faculty of Science, and 29.68% for the Faculty of Literature and Human Sciences.

4.3.2 Research design

The methodology of the study includes the following three (3) steps:

(1) *Translation into French of questionnaire, which were originally drafted in English*

CIP questionnaire is edited both in French and English. PIIA, modified PII, and RRPII questionnaire are only edited in English. The Table 14 presents the questionnaire of the four measurement models in English. All of the four English questionnaires can be found in the book of Bearden, Netemeyer and Haws (2011). Two experts translated PIIA, modified PII, and RRPII into French. The final French version is presented in Appendix 6.

Table 14. PIIA, modified PII, RRPII and CIP measurement models.

Items	1	2	3	4	1 : PIIA – 2 : modified PII – 3 : RRPII – 4 : CIP
x1	*	∞	∞		Important – Unimportant*
x2		∞	∞		Of no concern – Of concern to me
x3	*		∞		irrelevant – Relevant*
x4	*	∞	∞		Means a lot to me – Means nothing to me*
x5	*				Valuable – Worthless*
x6		∞	∞		Doesn't matter – Matters to me
x7		∞			Significant – Insignificant*
x8	+		~		Boring - Interesting
x9	+		~		Exciting – Unexciting*
x10	+		~		Appealing – Unappealing*
x11	+				Mundane – Fascinating
x12	*				Needed – Not needed
x13	+				Involving – Uninvolving*
x14			~		Fun – Not fun*
x15			~		Neat – Dull*
y1				○	If, after I bought _____, my choice(s) prove to be poor, I would be really upset
y2				○	When you choose _____, it is not a big deal if you make a mistake*
y3				○	It is really annoying to purchase _____ that are not suitable
y4				×	One can say _____ interests me a lot
y5				×	I attach great importance to _____

y6			×	_____ is a topic which leaves me totally indifferent*
y7			□	When one purchases ______, one is never certain of one's choice
y8			□	When I face a shelf of ______, I always feel a bit at a loss to make my choice
y9			□	Choosing ______ is rather complicated
y10			□	Whenever one buys ______, one never really knows whether they are the ones that should have been bought
y11			◊	The ______ you buy tells a little bit about you
y12			◊	The ______ I buy gives a glimpse of the type of man/woman I am
y13			◊	You can tell a lot about a person by the ______ he or she chooses
y14			●	______ is somewhat a pleasure to me
y15			●	Buying ______ is like buying a gift for myself
y16			●	It gives me pleasure to purchase ______

Legend:

◊ 'Sign value' dimension that the consumer assigns to the product

○ 'Perceived importance' dimension of negative consequences if the choice was mistaken

□ 'Subjective probability' dimension of making the wrong choice

× 'Personal interest' dimension in the product category

● Product's 'hedonic value' dimension

* 'Rational relevance' dimension

+ 'Emotional relevance' dimension

∞ 'Importance' dimension for the modified PII and RRPII

~ 'Interest' dimension for the RRPII

(2) *Data processing and analysis*

Processing of non-responses

Table in the Appendix 7 on univariate statistics, reveals a low (i.e. less than 5%) non-response rate for every item. Since the rate is less than 5% for every item, the impact on the data analysis is relatively weak (Kline, 2011). Little's MCAR (Little, 1988) test rejects the hypothesis that the missing data are not random at an alpha threshold of 5% (Chi-squared = 1717.793, DF = 1600, Sig. = 0.020). Therefore, ignoring these missing values will not skew the analysis, but might cause loss of precision. To avoid such loss, we applied the FIML imputation method. Our data set contains a total of 512 observations.

Testing multivariate normality of variables and selecting the estimation method

The next step is to test the multivariate normality of the variables. The Mardia test significantly rejects the null hypothesis of multivariate normality of the variables (multivariate skewness statistic = 463.464, p-value.skew < 0.010, multivariate kurtosis statistic = 2441.187, p-value.kurt < 0.010). This result requires using a robust estimation method, and we use MLR, which enables correcting estimated parameters and standard errors.

(3) *Comparing measurement models*

Two types of comparison are made in this study: intra-model and inter-model. Intra-model comparisons consist of the comparison of CFA and ESEM results through the psychometric criteria, which are reliability, convergent validity, discriminant validity, nomological validity, and goodness-of-fit. Inter-model comparisons are based on a synthesis of various tests performed for each model. Other pragmatic criteria are incorporated, such as the measurement model's parsimony and simplicity. Since the four measurement models are reflective in nature, the Jöreskog rho reliability test can be applied. Compared to the standard test using Cronbach's alpha, Jöreskog's rho is less sensitive to the number of items (Didellon & Valette-Florence, 1996). A Jöreskog rho value greater than 0.7 is conventionally deemed satisfactory proof that measurement models are reliable. For ESEM, the reliability test of a measurement model will not be applied due to the presence of crossed items supposed to measure other constructs. The criteria of Fornell and Larcker (1981) and those of Hair, Black, Babin, and Anderson (2010) will be applied for evaluating discriminant and convergent validity.[24] Nomological validity refers to the measurement model's ability to imitate the behavior supposedly displayed by the construct it operationalizes (Lynch, 1983). The importance of searching for information was chosen as the variable to be predicted, which is the only common variable for testing the nomological validity of the four measurement models in the course of their development (Figure 7).

The measurement models' goodness-of-fit to the data will be determined by means of the CFI, TLI, RMSEA, and SRMR measures. These measures are selected because they are relatively insensitive to sample size and can detect which model is poorly specified by comparing its performance to that of the other measures (Fan et al., 1999). The following criteria of Hu and Bentler (1999) will be applied: TLI ≥ 0.96 and SRMR ≤ 0.09, RMSEA ≤ 0.06 and SRMR ≤ 0.09, CFI ≥ 0.95 and SRMR ≤ 0.09, as well as the Chi-squared/df > 5 ratio as not to reject the hypothesis that a model fits the data.

Since the CFA is a constrained version the ESEM model, an unconstrained model can be tested against a constrained model to assess whether significant differences exist between the two approaches.

Although fit indices such as CFI, TLI, SRMR, and RMSEA are preferable for measuring the models' goodness-of-fit to the data, no studies have been done, to the best of our knowledge, on the performance of these measures when performing a comparison test for an unconstrained model compared to a constrained model. Some studies rely on fairly heuristic references (e.g., Cheung & Rensvold, 2002; Marsh et al., 2009). For TLI and CFI, for example, a variation of 0.01 indicates a difference between the two models being compared (Cheung & Rensvold, 2002). In the case of SRMR, variation must be greater than or equal to 0.001 (e.g. Marsh et al., 2009). The same applies to other information criterion measures, such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). These two criteria are seldom interpreted to evaluate consumer behaviour models in marketing research. To satisfy the criterion of parsimony, they penalize models that have too many parameters. On finding a significant difference between models, the model with the lower AIC or BIC is preferred. Conventionally, a difference of fewer

24 Indicator weights should be greater than 0.707 to prove that more than half of the variance is captured by the latent construct (Gefen, Straub, & Boudreau, 2000). Nonetheless weights between 0.500 and 0.700 are acceptable (Hair et al. , 2010).

than 10 points between the two models being compared favours the model with the smaller AIC (Burnham & Anderson, 2002). Similarly, for the BIC, a difference of minus 10 points between unconstrained and constrained models favours the more complex model (Raftery, 1993).

To obtain an objective statistical criterion not based on heuristic methods, one can use a classical and robust[25] differentiation method, namely the Chi-squared test. This test allows us to decide whether a significant difference exists between an unconstrained and its constrained model by relying on a statistical threshold (e.g. alpha = 1%). As such, the dimensional structure of different measurement models can also be compared by using this test. For comparing dimensional structure, the tests consist in comparing constrained models with an unconstrained model, such as comparing a one-dimensional with a two-dimensional structure of the RRPII model. In this example, the one-dimensional model is the constrained version of the two-dimensional model, regardless of whether the test uses the CFA or ESEM approaches.

The ESEM model will be estimated by means of the geomin rotation method with a value $\varepsilon = 0.0001$ (Asparouhov & Muthén, 2009). The geomin method was selected both to minimize the complexity factor[26] and ease interpretation of factor contributions. To analyse ESEM, the item factor contributions are examined in relation to their ability to accurately measure the variables for which they are intended. Cross-factor contributions larger than the contribution that measures the factor represent a poor specification of the model. Consequently, we selected a threshold of significance of $p < 0.05$. However, we consider factor contributions important only if they exceed the value of 0.5 recommended by Hair et al. (2010).

Lastly, the cross-validation method will be applied to evaluate the invariance of the models' results by using different samples of the same population (in our case, the students of the University of Neuchâtel). The test to be considered for the hypothesis is the following:

[25] MLR robust estimation method.
[26] That is, weak cross-factor contributions and strong inter-factor correlations.

H0: The parameters (weighting factors, variances and covariances of the factors, measurement errors) of the measurement model being tested are invariant between two samples of the same population.

Against

H1: At least one of the measurement model's parameters is not identical between the two samples.

Figure 7. Models for testing nomological validity.

Figure 7a. CIP model for the CFA

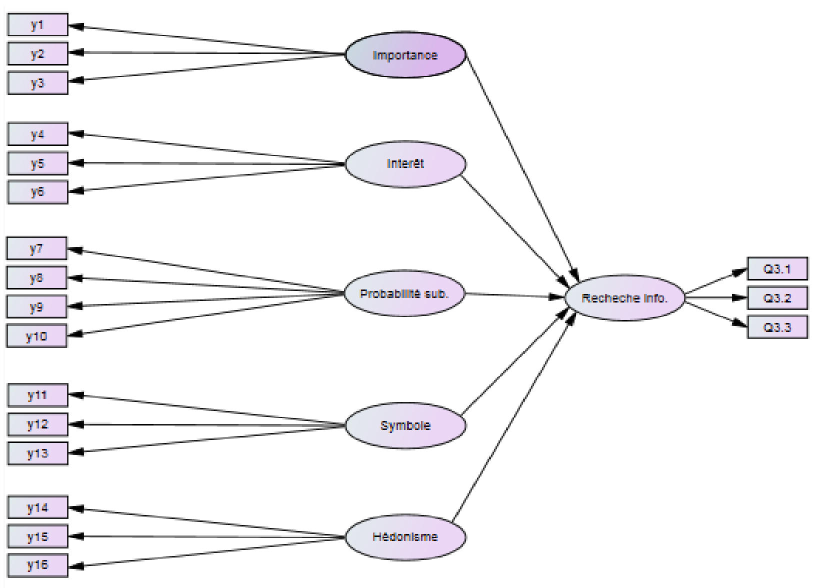


Figure 7b. CIP model for the ESEM

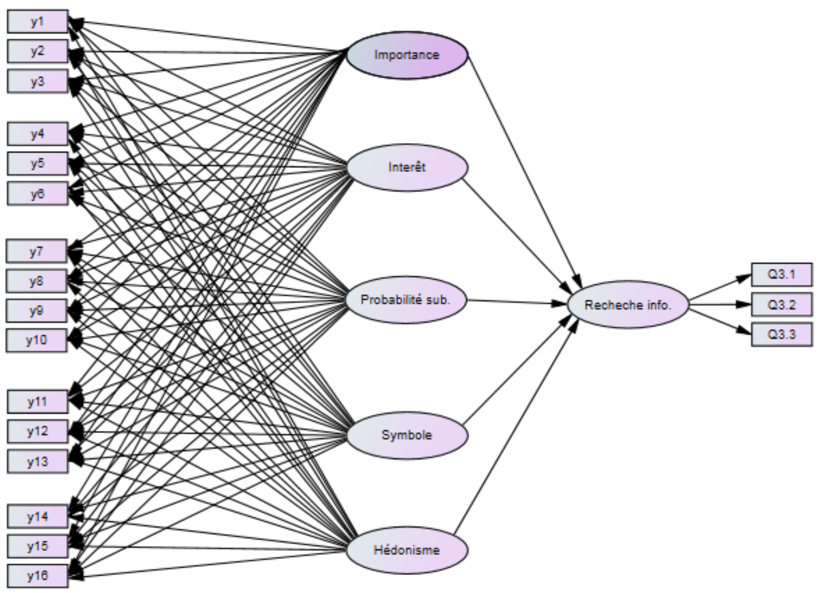


Figure 7c. Modified PII model

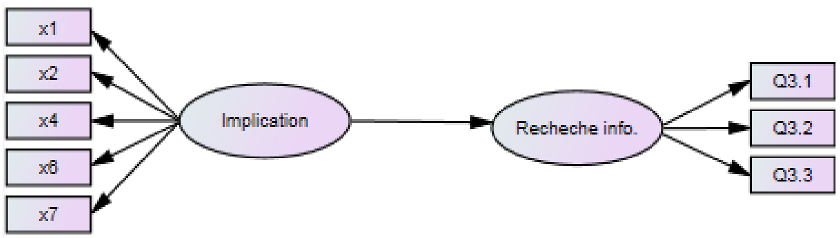


Figure 7d. One-dimensional PIIA model

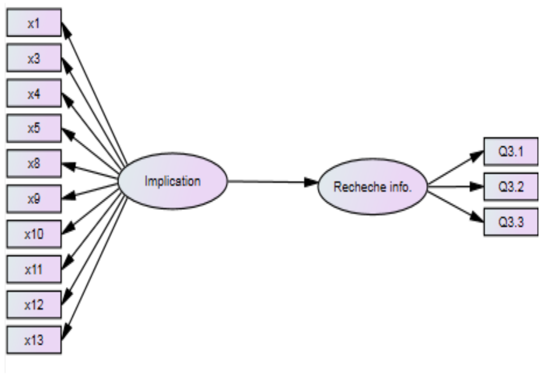


Figure 7e. Two-dimensional PIIA model for the CFA

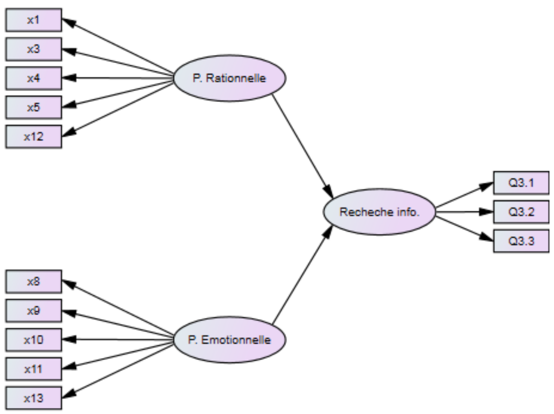


Figure 7f. Two-dimensional PIIA model for the ESEM

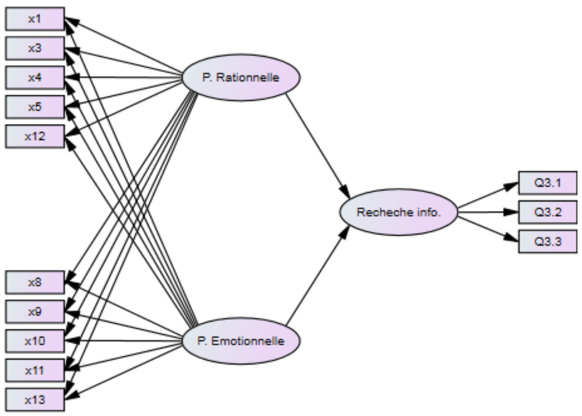


Figure 7g. One dimensional RRPII model

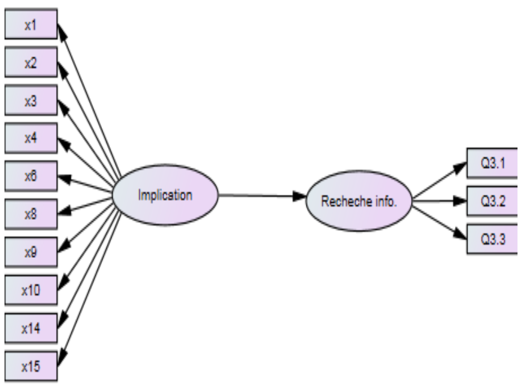


Figure 7h.Two-dimensional RRPII model for the CFA

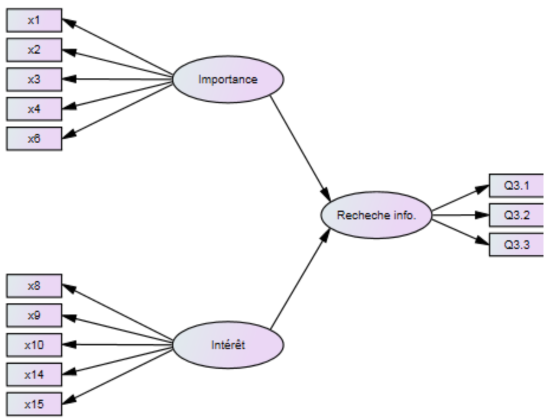
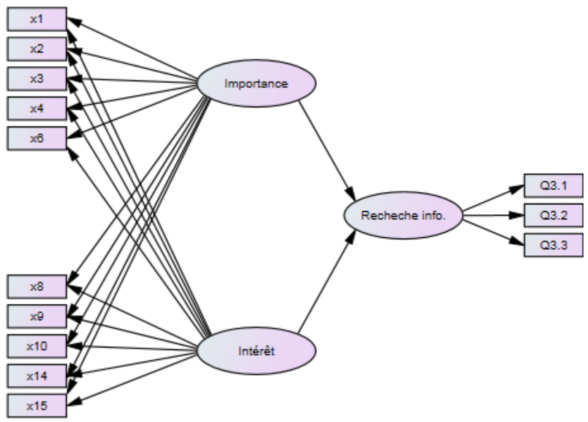


Figure 7i.Two-dimensional RRPII model for the ESEM



4.4 RESULTS

Before beginning to analyze the results, we present below the results of evaluating the quality of and differences in the measurement models, depending on the approach chosen (CFA or ESEM). Table 15 presents the results for the different measures of the models' fitness to the data, sorted by method and the model's dimensional structure. Then, Table 16 shows the results of comparing models in terms of dimensional structure and method. Finally, Table 17 presents the results for nomological validation across measures for the different models.

Table 15. Results for data fit measures, sorted by the method and structural dimension of the models.

Model description		χ^{2*}	*df*	χ^2/df^*	*Correction factor*	*Loglikelihood (MLR)*	*Parameter*	*Correction factor*	*AIC*	*BIC*	*RMSEA*	*Confidence interval for RMSEA 90%*	*CFI*	*TLI*	*SRMR*
CIP	*CFA*														
	Five dimensions	192.252	94	2.080	1.210	-10939.100	58	1.205	21994	22240	0,045	0,036 ; 0,054	0,955	0,943	0,042
	ESEM														
	Five dimensions	101.524	50	2.034	1.111	-10879.143	102	1.256	21962	22394	0,045	0,032 ; 0,057	0,977	0,944	0,021
PII	*CFA*														
	One dimension	201.328	35	5.979	1.186	-7142.709	30	1.079	14345	14472	0,096	0,084 ; 0,109	0,857	0,816	0,057
	Five dimensions	156.632	34	4.817	1.177	-7115.544	31	1.092	14293	14424	0,084	0,071 ; 0,097	0,895	0,861	0,055
	ESEM														
	Two dimensions	168.226	26	5.616	0.803	-7090.922	39	1.358	14259	14425	0,103	0,089 ; 0,119	0,878	0,789	0,039
Modified PII	*CFA/ESEM*														
	One dimension	25.773	5	5.110	1.294	-3658.352	15	1.108	7346	7410	0,090	0,058 ; 0,126	0,950	0,888	0,036
RRPII	*CFA*														
	One dimension	256.3961	35	7.638	1.249	-7020.585	30	1.090	14101	14228	0,111	0,099 ; 0,124	0,862	0,822	0,069
	Two dimensions	58.277	34	4.840	1.221	-6956.989	31	1.133	13975	14024	0,084	0,071 ; 0,098	0,922	0,897	0,055
	ESEM														
	Two dimensions	61.381	26	2.313	1.161	-6896.011	39	1.191	13870	14035	0,052	0,035 ; 0,068	0,978	0,962	0,024

Table 16. Results for comparisons between models, sorted by method and by structural dimension.

Model comparison		$\Delta\chi^{2*}$	Δdf	$\chi^{2*}/\Delta df$	*-2*ΔLog MLR*	*-2*ΔLog MLR/Δdf*	*ΔAIC*	*ΔBIC*	*ΔRMSEA*	*ΔCFI*	*ΔTLI*	*ΔSRMR*
CIP	*Full model : CIP with ESEM* *Against* *Nested model : CIP with CFA*	90.610	44	2.059	90.622	2.059	-31.914	154,572	0.000	0,022	0,001	- 0,021
PII	*Full model : PII (2 dimensions) with ESEM* *Against* *Nested model : PII (1 dimension) with ESEM*	45.184	9	5.022	45.268	5.029	-85,574	-47,428	0,007	0,021	-0,027	-0,018
	Full model : PII (2 dimensions) with CFA *Against* *Nested model : PII (1 dimension) with CFA*	36.474	1	36.474	36.660	36.660	-52,33	-48,091	-0,012	0,038	0,045	-0,002
	Full model : PII (2 dimensions) with ESEM *Against* *Nested model : PII (2 dimensions) with CFA*	20.594	8	2.574	20.615	2.576	-105,956	11,096	-0,032	0,056	0,065	-0,031
Modified PII	*CFA and ESEM are similar*	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
RRPII	*Full model : RRPII (2 dimensions) with ESEM* *Against* *Nested model : RRPII (1 dimension) with ESEM*	165.628	9	18.403	163.091	18.121	-231,148	-193,003	-0,059	0,116	0,140	-0,045
	Full model : RRPII (2 dimensions) with CFA *Against* *Nested model : RRPII (1 dimension) with CFA*	57.693	1	57.693	52.494	52.494	-125,192	-252,341	-0,027	0,06	0,075	-0,014
	Full model : RRPII (2 dimensions) with ESEM *Against* *Nested model : RRPII (2 dimensions) with CFA*	86.153	9	9.572	86.142	9.571	-105,956	11,096	-0,032	0,056	0,065	-0,031

Table 17. Test results for the nomological model across measures.

Nomological validity		χ^{2*}	*df*	χ^2/df^*	*Correction factor*	*Loglikelihood (MLR)*	*Paramenter*	*Correction factor*	*AIC*	*BIC*	*RMSEA*	*Confidence interval for RMSEA 90%*	*CFI*	*TLI*	*SRMR*
CIP	*CFA*														
	Five dimensions	277,660	137	2.026	1,167	-12899,146	72	1,167	25942	26247	0,045	0,037 ; 0,052	0,956	0,945	0,041
	ESEM														
	Five dimensions	174,941	93	1.881	1,107	-12833,896	116	1,215	25899	26391	0,041	0,032 ; 0,051	0,974	0,953	0,024
PII	*CFA*														
	One dimension	252,525	64	3.945	1,154	-9152,673	40	1,056	18385	18554	0,076	0,066 ; 0,086	0,905	0,885	0,051
	Two dimensions	204,229	62	3.293	1,149	-9124,239	42	1,068	18332	18510	0,067	0,057 ; 0,077	0,929	0,910	0,048
	ESEM														
	Two dimensions	250,796	54	4.644	0,721	-9097,427	50	1,543	18294	18506	0,084	0,074 ; 0,095	0,901	0,857	0,036
Modified PII	*CFA*														
	One dimension	71,586	19	3.767	1,178	-5685,567	25	1,075	11421	11527	0,074	0,056 ; 0,092	0,953	0,931	0,042
RRPII	*CFA*														
	Two dimensions	202,570	62	3.267	1,183	-8955,396	42	1,107	17994	18172	0,067	0,056 ; 0,077	0,943	0,928	0,051
	ESEM														
	Two dimensions	97,102	54	1.798	1,145	-8891,117	50	1,160	17882	18094	0,039	0,026 ; 0,052	0,982	0,975	0,026

4.4.1 Analysis of intra-model results

CIP measurement model

On the basis of Hu and Bentler's (1999) criteria, used in this study, the results for the alternative measures do not reject the suitability of either of the two approaches – CFA and ESEM – for the CIP model (see Table 15).

Regarding the internal reliability test, Jöreskog's internal consistency coefficient rho of the five dimensions of the CIP seem satisfactory except for the 'perceived importance' dimension, which has a value of 0.670. Nevertheless, it approximates the conventional value of 0.7.

Analysis of each dimension's convergent validity indicates in both cases that:

- the weights are significantly different from zero (see Table 18). By contrast, in terms of their values, 10 of the 16 items are satisfactory (weights > 0.707) for CFA and 8 out of 16 items for the ESEM. On the other hand, item y8 poses a problem both for CFA and ESEM.
- r^2 for items y2 and y3 of the 'perceived importance' dimension, y8 and y9 of the 'subjective probability', dimension and y13 of the 'sign value' dimension are below the conventional threshold of 0.5 (see Table 18)
- AVEs for the 'perceived importance' (AVE = 0.418) and the 'subjective probability' dimensions (AVE = 0.425) are relatively close to 0.5 (see Table 19)

The square roots of the AVE of each dimension are generally greater than the correlations shared with the other constructs, except that of 'interest' (AVE = 0.770), which is below the correlation between the 'interest' and 'hedonism' dimensions (r = 0.793) for CFA (see Table 19). Nonetheless, considering the value of the correlation yielded by the ESEM (r = 0.639), discriminant validity is demonstrated. However, the ESEM detects a cross-weighting problem for the y14 and y9 indicators, which measure other constructs, namely 'interest' and 'subjective probability' (see Table 18). For the test of the nomological model only the 'perceived importance' and 'subjective probability' dimensions have coefficients that are not significant at an alpha threshold of 0.1 for the CFA model (See Table 17). For ESEM, the 'hedonic' dimension also becomes non-significant. Fit indices indicate the data's overall goodness-of-fit to acceptable models.

Table 18. Results for CIP factor weights.

CIP dimensions	Obs.	*1st factorial axis*		*2nd factorial axis*		*3rd factorial axis*		*4th factorial axis*		*5th factorial axis*		*R-squarred*	
		CFA	***ESEM***	***CFA***	***ESEM***	***CFA***	***ESEM***	***CFA***	***ESEM***	***CFA***	***ESEM***	***CFA***	***ESEM***
Perceived importance	Y1	0.796*	0.645*	-	0.100	-	0.120+	-	0.013	-	0.067	0.634	0.546
	Y2	0.651*	0.630*	-	0.103	-	-0.029	-	0.078	-	-0.053	0.423	0.471
	Y3	0.444*	0.556*	-	-0.095	-	-0.059	-	-0.077	-	0.059	0.197	0.273
Interest	Y4	-	-0.042	0.737*	0.793*	-	-0.082	-	0.001	-	-0.029	0.542	0.601
	Y5	-	0.089	0.761*	0.644*	-	0.006	-	0.020	-	0.087	0.579	0.568
	Y6	-	0.063	0.809*	0.724*	-	0.031	-	0.000	-	0.079	0.655	0.643
Subjective probability	Y7	-	-0.001	-	0.019	0.730*	0.707*	-	0.024	-	-0.024	0.533	0.502
	Y8	-	0.139+	-	-0.026	0.462*	0.406*	-	-0.059	-	-0.121	0.214	0.223
	Y9	-	0.251*	-	-0.020	0.531*	0.480*	-	0.010	-	-0.098	0.282	0.321
	Y10	-	-0.044	-	-0.009	0.821*	0.895*	-	-0.004	-	0.143	0.674	0.774
Sign-value	Y11	-	0.004	-	0.004	-	-0.022	0.780*	0.799*	-	0.009	0.608	0.649
	Y12	-	0.046	-	-0.030	-	-0.016	0.772*	0.708*	-	0.115	0.596	0.560
	Y13	-	-0.039	-	0.095	-	0.054	0.612*	0.604*	-	-0.056	0.374	0.389
Hedonism	Y14	-	-0.002	-	0.269*	-	-0.022	-	0.036	0.850*	0.608*	0.725	0.682
	Y15	-	0.118+	-	-0.023	-	-0.001	-	0.067	0.695*	0.658*	0.483	0.511
	Y16	-	-0.008	-	0.117	-	0.005	-	-0.043	0.837*	0.818*	0.701	0.774

The mean of the R-squared is 0.513 for the CFA and 0.530 for the ESEM

Table 19. Results for CIP inter-factor correlations.

CIP interfactor correlation	*Perceived importance*		*Interest*		*Subjective probability*		*Sign value*		*Hedonism*	
	CFA	*ESEM*	*CFA*	*ESEM*	*CFA*	*ESEM*	*CFA*	*ESEM*	*CFA*	*ESEM*
Perceived importance	**0.646**	1	-	-	-	-	-	-	-	-
Interest	0.544	0.414	**0.770**	1	-	-	-	-	-	-
Subjective probability	0.164	0.067	-0.107	-0.142	**0.652**	1	-	-	-	-
Sign value	0.353	0.262	0.515	0.462	-0.007	-0.019	**0.725**	1	-	-
Hedonism	0.459	0.323	0.793	0.639	-0.117	-0.158	0.429	0.309	**0.797**	1

The PIIA measurement model

Applying the reliability test to both one- and two-dimensional PIIA models yielded acceptable rhos for both models, although for the ‘rational relevance’ dimension the value of rho is 0.683, which is slightly below the conventional threshold of 0.7.

Statistical tests reveal a discrepancy between the data gathered and models chosen, either one- or two-dimensional (see Table 15). The fit indices do not comply with the criteria recommended by Hu and Bentler (1999).

If we compare the two models (i.e. constrained and unconstrained) of the CFA, and those of the ESEM models, we find a significant difference between the one-dimensional and the two-dimensional PIIA models (see Table 16).

Examination of the one-dimensional structure’s convergent validity shows that:

- based on analysis of the item weights, the entire acceptable convergent validity exists except for the three items x3, x12, and x13, whose values are less than 0.5 (see Table 20),
- all r^2 are less than 0.5 except for x10 (see Table 20),
- AVE are equal to 0.338 (less than the threshold 0.5).

In view of these results, one can conclude lack of convergent validity of the one-dimensional structure. Analysis of the convergent validity of the two-dimensional structure underlines a discrepancy between the results for CFA and ESEM with respect to the ‘rational relevance’ dimension (Table 21). For CFA, the weights are significantly non-zero, while for ESEM, items x4, x5, and x12 are not significant. By contrast, both methods are conclusive with respect to the lack of convergence of the indicator x12 on the basis of its weighting coefficient and the value of coefficient r2. For the ‘emotional relevance’ structure, both methods yield similar results, supporting the convergence of items, except for x13 that has a light weight (> 0.5) and a fairly heavy weight of 0.633 for the ‘rational relevance’ dimension it is not even supposed to measure. The AVEs of both sub-dimensions are above the recommended threshold of 0.5 (Table 22).

Regarding discriminant validity, indicator x13 is problematic, yielding a fairly heavy weight on the ‘rational relevance’ dimension (0.633), which it is not supposed to measure, and a light weight on the ‘emotional relevance’ dimension (0.007) to which it is attributed. Overall, the ESEM method approves the model, unlike the CFA method. In this situation, the results for the CFA method must be interpreted with caution, given the method’s limits, where it overestimates correlation coefficients due to imposing zero constraints (Marsh et al., 2011, 2010, Cole et al., 2007).

For the test of nomological validity, the one-dimensional structure is nomologically validated. For the two-dimensional model, we obtain the same results as Zaichkowsky (1994), whereas we had expected, first, a stronger link between the cognitive dimension and the search for information and, second, less linkage to the affective dimension. Nonetheless, Zaichkowsky (1994) findings are supported by those of this paper, since the strongest links are those between the affective dimension and the search for information.

Table 20. Results for factor weights for the one-dimensional PII.

PII dimensions	*Obs.*	*1st factorial axis*		*R-squarred*	
		CFA	***ESEM***	***CFA***	***ESEM***
Rational relevance	***X1***	0.589*	0.589*	0.347	0.347
	X3	0.458*	0.458*	0.209	0.209
	X4	0.625*	0.625*	0.390	0.390
	X5	0.654*	0.654*	0.428	0.428
	X12	0.180*	0.180*	0.032	0.032
Emotional relevance	***X8***	0.607*	0.607*	0.368	0.368
	X9	0.650*	0.650*	0.429	0.429
	X10	0.753*	0.753*	0.568	0.568
	X11	0.594*	0.594*	0.353	0.353
	X13	0.513*	0.513*	0.263	0.263

The mean of the R-squared is 0.338 for the CFA and the ESEM

Table 21. Results for factor weights for the two-dimensional PII.

PII dimensions	*Obs.*	*1st factorial axis*		*2nd factorial axis*		*R-squarred*	
		CFA	***ESEM***	***CFA***	***ESEM***	***CFA***	***ESEM***
Rational relevance	***X1***	0.615*	0.457+	-	0.224	0.378	0.387
	X3	0.469*	0.666*	-	-0.077	0.220	0.384
	X4	0.700*	0.428	-	0.274	0.490	0.406
	X5	0.725*	0.369	-	0.355	0.525	0.427
	X12	0.184*	0.145	-	0.061	0.034+	0.036
Emotional relevance	***X8***	-	-0.090	0.646*	0.720*	0.417	0.445
	X9	-	0.027	0.688*	0.663*	0.474	0.462
	X10	-	-0.002	0.792*	0.806*	0.627	0.648
	X11	-	0.070	0.604*	0.557*	0.365	0.365
	X13	-	0.633*	0.477*	0.007	0.227	0.407

The mean of the R-squared is 0.413 for the CFA and 0.396 for the ESEM

Table 22. Results for inter-factor correlations for the two-dimensional PII.

PII (2 dim.) corrélation inter-factorielle	*Rational relevance*		*Emotional relevance*	
	CFA	*ESEM*	*CFA*	*ESEM*
Rational relevance	**0.700**	1	-	-
Emotional relevance	0.809	0.629	**0.750**	1

The modified PII measurement model

The combined CFI and SRMR criterion of Hu and Bentler (1999)[27] indicates that the model is consistent with the data (see Table 15)

The rho reliability coefficient of the measurement model is satisfactory (rho = 0.749). The results for the weighting coefficients indicate acceptable and significant coefficients (see Table 23)

However, the AVE remains relatively low (AVE = 0.376) compared to the recommended threshold for the convergence criterion. Due to its one-dimensional structure, discriminant validity was not evaluated. On testing its nomological validity, the model was validated for an alpha threshold of 0.1.

Table 23. Results for factor weights for the modified PII model (Mittal, 1995).

Modified PII dimension	***Obs.***	***1st factorial axis***		***R-squarred***	
		CFA	***ESEM***	***CFA***	***ESEM***
Importance	***X1***	0.566*	0.566*	0.321	0.321
	X2	0.632*	0.632*	0.400	0.400
	X4	0.722*	0.722*	0.522	0.522
	X6	0.596*	0.596*	0.355	0.355
	X7	0.536*	0.536*	0.288	0.288

[27] CFI ≥ 0.95 and SRMR ≤ 0.09

The RRPII measurement model

For both models (one-dimensional and two-dimensional), the reliability threshold of 0.7 was reached. As for the model's goodness-of-fit to the data, the two-dimensional model fits the data better than the one-dimensional one (Table 15). Nonetheless, unlike the two-dimensional ESEM model, its CFA counterpart does not comply with the selected threshold criteria (Table 15).

The statistical evaluation for comparing the two models reveals a significant difference between the one-dimensional and two-dimensional models (see Table 15). This supports hypothesis H0d on dimensional structure, according to which the RRPII is two-dimensional. Therefore analysing the psychometric qualities of the one-dimensional structure is not useful.

However, although the results for CFA and ESEM models converge in terms of the RRPII dimensional structure, when comparing the results for these two methods for the two-dimensional structure, a significant difference in statistical results is found (see Table 16). Moreover, unlike CFA, the ESEM model yields acceptable results with reference to the criteria of Hu and Bentler (1999).

Regarding analysis of convergent validity, the 'importance' dimension has an AVE equal to 0.360 and the 'interest' dimension an AVE equal to 0.594 (see Table 26). Nonetheless, in terms of prediction, the r^2 of certain items (particularly the 'importance' dimension) are below 0.5 both for the CFA and the ESEM models.

For analysis of the discriminant validity, examination of the weighting factors of the crossed items shows that items x8, x10, and x2 are particularly problematic (see Table 25). The first two, which should measure only the 'interest' construct, have significant weighting coefficients with a different construct that is not supposed to be measured, namely 'importance.' The same holds true for the third factor, which should, as a rule, measure the 'importance' construct and yield a significant weighting factor with the 'interest' construct. Analysis of the correlation matrix reveals that there is specifically a problem of discriminant validity in the 'importance' dimension (AVE = 0.6, corr (Importance, interest) = 0.745 for CFA and corr (Importance, Interest) = 0.63 for ESEM).

Additionally, the nomological validity test was validated for an alpha threshold of 0.1. The fit indices generally indicate a satisfactory fit of the data to the two-dimensional RRPII model.

Table 24. Results for factor weights for the one-dimensional RRPII.

RRPII dimensions	***Obs.***	***1st factorial axis***		***R-squarred***	
		CFA	***ESEM***	***CFA***	***ESEM***
Importance	***X1***	0.467*	0.467*	0.210	0.210
	X2	0.632*	0.632*	0.399	0.399
	X3	0.345*	0.345*	0.120	0.120
	X4	0.518*	0.518*	0.268	0.268
	X6	0.498*	0.498*	0.248	0.248
Interest	***X8***	0.656*	0.656*	0.430	0.430
	X9	0.826*	0.826*	0.682	0.682
	X10	0.712*	0.712*	0.507	0.507
	X14	0.767*	0.767*	0.588	0.588
	X15	0.840*	0.840*	0.706	0.706

The mean of the R-squared is 0.415 for the CFA and the ESEM

Table 25. Results for factor weights for the two-dimensional RRPII.

RRPII dimensions	***Obs.***	***1st factorial axis***		***2nd factorial axis***		***R-squarred***	
		CFA	***ESEM***	***CFA***	***ESEM***	***CFA***	***ESEM***
Importance	***X1***	0.595*	0.709*	-	-0.104	0.354	0.421
	X2	0.699*	0.532*	-	0.213*	0.489	0.471
	X3	0.433*	0.447*	-	-0.011	0.187	0.193
	X4	0.636*	0.544*	-	0.090	0.405	0.366
	X6	0.606*	0.627*	-	0.002	0.368	0.395
Interest	***X8***	-	0.309*	0.641*	0.414*	0.411	0.428
	X9	-	0.006	0.850*	0.855*	0.723	0.738
	X10	-	0.524*	0.676*	0.302*	0.457	0.565
	X14	-	0.010	0.787*	0.782*	0.616	0.621
	X15	-	0.027	0.872*	0.909*	0.760	0.797

The mean of the R-squared is 0.477 for the CFA and 0.499 for the ESEM

Table 26. Results for factor correlations for the two-dimensional RRPII.

RRPII (2 dim.) corrélation inter-factorielle	***Importance***		***Interest***	
	CFA	***ESEM***	***CFA***	***ESEM***
Importance	**0.600**	1	-	-
interest	0.745	0.630	**0.770**	1

4.4.2 Analysis of inter-modal results

The evaluation of parsimony depends on the number of items. The shorter the questionnaire, the lower the non-response rate (individuals who did not check a box for a given item). In this case, the modified PII comes first with five items. The RRPII and PIIA (each consisting of 10 items) are equal and hold second place. CIP, due to its 16 items, is last.

In terms of simplicity, the differential semantic scale (PIIA, RRPII, modified PII) has an advantage over the Likert (CIP) scale in terms of information content. A single declarative sentence might suffice to apply all items to a particular product. This is very convenient for paper or online questionnaires. One advantage of using the differential semantic scale instead of the differential scale is that it is simple to run for evaluating the intensity and meaning that the respondent assigns to the concept being measured (Mindak, 1961).

However, this benefit should be used with caution. Since CIP items consist of complete sentences, they are less likely than differential semantics to be affected by systematic responses.[28] Moreover, the Likert scale is more practical than the differential semantic scale when administering the questionnaire by mail, telephone, or face-to-face interview (Malhotra et al., 2011). In short, the advantage in terms of ease of use is shared, since it depends on the manner of administering the questionnaires. As for the analysis of internal consistency (reliability analysis), no particular obstacles to internal consistency were noted in the measurement models. Only the 'rational relevance' dimension of the PIIA and the 'perceived importance' dimension of the CIP have a Jöreskog rho value slightly below 0.7.

For analysing convergent validity, the first criterion of significance was overall complied with by all the measurement models. With regard to the criterion stating that the AVE must be greater than or equal to 0.5, the dimensions of the following measurement models failed to attain the recommended threshold:

- "rational relevance" of the PIIA model with an AVE of 0.490;
- "importance" of the modified PII model with an AVE equal to 0.376;
- "importance" of the RRPII model with an AVE equal to 0.360;

[28] Note that the differential semantic scale is still relatively insensitive to systematic response errors (Friborg et al., 2006).

- “perceived importance” and “subjective probability” of the CIP model, with AVEs respectively of 0.418 and 0.425.

Overall CIP and PIIA perform better, since the AVE values of ‘rational relevance,’ ‘perceived importance,’ and ‘subjective probability’ are relatively close to the threshold criterion of 0.5. With respect to the prediction criterion r^2, its average for the items and the percentage of items with a r^2 of less than 0.5 were calculated for each measurement model, for both CFA and ESEM. The CIP comes first in terms of overall performance with satisfactory r^2 values on average, namely 0.513 for CFA and 0.530 for the ESEM. The percentage of items beneath 0.5 is 37.5% for CFA and 31.25% for ESEM. In second place, we find the RRPII, with mean r^2 values close to the threshold: 0.477 for CFA and 0.499 for the ESEM. The percentage of items beneath the recommended threshold is 70% for CFA and 60% for ESEM. In third and fourth place, respectively, are the PII and the modified PII with average r^2 values relatively close to each other, namely 0.413 for CFA and 0.396 for ESEM for the first and 0.377 for CFA and the ESEM for the second. Of these three converging validity criteria, the CIP comes first in terms of performance, followed by both the RRPII and PIIA scales. RRPII is better than PIIA on r^2 criteria but not on AVE. Modified PII holds third place.

For the discriminant validity analysis, the results diverge depending on the method chosen for the PIIA and CIP measurement models. The ESEM method confirms discriminant validity, unlike the CFA method. For RRPII, discriminant validity could not be confirmed. In sum, we propose the following ranking: PIIA and CIP hold first place, followed by the RRPII. For the modified PII model, no test was carried out because of its one-dimensional structure. For nomological validity, only the ‘perceived importance’ and ‘subjective probability’ dimensions of the CIP and the ‘rational relevance’ dimension of the PIIA were not validated nomologically. For the remaining items, the criterion of validity was validated.

Finally, with regard to the models’ goodness of fit to the data, we can rank the four scales in order of their compliance with the recommendations of Hu and Bentler (1999). First, CIP has good fit for both CFA and ESEM. Second comes RRPII, with good results only when using the ESEM method. Third is the modified PII with mixed results between, on one hand, the group consisting of the CFI and SRMR measures, which do not reject the model, and on the other hand, the group with measures TLI ≥ 0.96 and SRMR ≤ 0.09 and RMSEA ≤ 0.06 and SRMR ≤ 0.09, which reject the model. Last is PIIA, whose measures are outside the recommended ranges. Table 27 shows a synopsis of this ranking.

Table 27. Synopsis for the comparison of the four measurement models.

Ranking	Parsimony	Simplicity	Reliability	Convergent validity	Discriminant validity	Nomological validity	Goodness-of-fit
1	Modified PII	Depending on the administration of the survey questionnaire	RRPII and Modified PII	CIP	PIIA et CIP	Modified PII and RRPII	CIP
2	RRPII and PIIA		CIP and PIIA	PIIA and RRPII	RRPII	CIP and PIIA	RRPII
3	CIP		-	Modified PII	-	-	Modified PII
4	-		-	-	-	-	PIIA

4.4.3 Cross-validation results for each of the four measurement

Table 28 below shows the results for cross-validation applied to each measurement model. We conclude that, starting from an alpha significance threshold of 1%, only the RRPII model rejects the null hypothesis. Nonetheless, with a strongly significant threshold of 0.1%, the null hypothesis is not rejected. In view of the weakness of the Chi-squared test, which rejects the null hypothesis more readily when the sample size is large (n > 200) (Schumacker & Lomax, 2010), the results obtained on the non-rejection of the null hypothesis can be considered reliable. The result of the Chi-squared/df index is below 5 for all models, likewise supporting the null hypothesis.

Table 28. Results for the difference test in Chi-squared.

PII	Constraint model	Chi-Square = 245.15, df=89
	Non constraint model	Chi-Square = 219.27, df=73
	Chi-Square Difference Tests	Chi-Square = 25.88, df = 16, p-value = 0.055
Modified PII	Constraint model	Chi-Square = 54.02, df=20
	Non constraint model	Chi-Square = 34.46, df=11
	Chi-Square Difference Tests	Chi-Square = 19.56, df = 9, p-value = 0.020
RRPII	Constraint model	Chi-Square = 268.93, df=89
	Non constraint model	Chi-Square = 225.16, df=70
	Chi-Square Difference Tests	Chi-Square = 43.77, df = 19, p-value = 0.001
CIP	Constraint model	Chi-Square = 279.91, df=200
	Non constraint model	Chi-Square = 256.32, df=170
	Chi-Square Difference Tests	Chi-Square = 23.59, df = 30, p-value = 0.790

4.5 DISCUSSION AND CONCLUSIONS

This study shows the results of comparing two different approaches to SEM, namely ESEM and CFA, and is designed to evaluate the psychometric qualities of both measurement models. To the best of our knowledge, no study of operationalization of constructs by means of the ESEM method has been proposed, within the discipline of consumer behavior research in marketing. ESEM combines the benefits of CFA and EFA. ESEM and CFA were evaluated by comparing four models for measuring consumer involvement (CIP, PIIA, modified PII, and RRPII). The involvement concept was chosen because of its importance in explaining consumers' purchasing behavior, and because of the controversy in the literature over the dimensional structure of its measurement models. Overall, the ESEM results diverged from the CFA results regarding the evaluation of discriminant validation for the CIP and PIIA measurement models. Unlike CFA, ESEM does not reject the discriminant validity of both measurement models.

Due to EFA's particularity regarding cross-weights, the ESEM method has the advantage over CFA of providing significantly more information on indicators likely to pose problems with convergent and discriminant validity. The application of ESEM has revealed indicators that measure other constructs they were not supposed to measure, such as two indicators of the CIP, three indicators of the two-dimensional RRPII, and one indicator of the two-dimensional PII. Consequently, either the translation must be re-checked or we must look for what significance the consumer attaches to these items. This characteristic of ESEM might help researchers improve the discriminant validity exhibited by measurement models. For example, after finding that the discriminant validity of the two-dimensional RRPII model was not validated, items three items were removed. The outcome was that discriminant validity was accepted. Another result that should be stressed is the RRPII model's goodness-of-fit to the data. The ESEM and CFA assessments contradict each other on this point: unlike CFA, the ESEM supports the RRPII model.

Regarding the evaluation of the nomological validation of the models, only the case of the CIP model shows a difference between the results of CFA and ESEM. The ESEM model yields a statistically insignificant coefficient for the 'hedonism' construct, in contrast to the CFA model. One explanation might be the presence of a multicollinearity problem in the CFA nomological model. CFA is likely to present problems of multicollinearity that affect structural parameters, in contrast to ESEM (Marsh et al., 2011, Asparouhov & Muthén, 2009, Cole et al., 2007). Moreover, the correlation matrix (see Table 19) indicates fairly large correlations (above 0.5) for some variables.

Finally, none of the measurement models was ranked the best on all the psychometric and pragmatic criteria. CIP seems to rank relatively well for convergent validity, discriminant validity, and goodness of fit. It is not well-ranked for parsimony, but fortunately the questionnaire is not too long (only 16 items). RRPII seems to be the second-best measurement model, and ranked second after CIP on average. Regarding PIIA, the results of the goodness of fit evaluations seem problematic. This makes the statistical results of all the psychometric criteria not relevant for PIIA.

Therefore, we classified the modified PII as third. However, the modified PII discriminant validity wasn't assessed because of its uni-dimensional structure. Finally, based on the data we have and the product we chose (footwear), the CIP and RRPII models seem to be better than the PIIA and the modified PII regarding psychometric criteria.

In conclusion, this study illustrates a difference in results from ESEM and CFA. Nevertheless, we should stress some limitations of the application of ESEM in this study. With ESEM we do not have a computation of the AVE or an internal consistency analysis. Therefore, we have to rely on the result obtained by CFA. In this regard, as long as AVE is utilized to assess convergent validity, ESEM cannot completely replace CFA, but should instead be treated as a complementary tool.

4.6 REFERENCES

Andrews, J., Durvasula, S., & Akhter, S. (1990). A Framework for Conceptualizing and Measuring the Involvement Construct in Advertising Research. *Journal Of Advertising*, 19(4), 27-40.

Asparouhov, T., & Muthén, B. (2009). Exploratory Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 397-438.

Bearden, W., Netemeyer, R., & Haws, K. (2011). *Handbook of marketing scales* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Bollen, K. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K., & Pearl, J. (2013). *Eight myths about causality and structural equation models in Handbook of causal analysis for social science* (pp. 301 - 330). New York: Springer.

Brannick, M. (1995). Critical comments on applying covariance structure modeling. *Journal Of Organizational Behavior*, 16(3), 201-213.

Browne, M. (2001). An Overview of Analytic Rotation in Exploratory Factor Analysis. *Multivariate Behavioral Research*, 36(1), 111-150.

Burnham, K., & Anderson, D. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: NY: Springer-Verlag.

Cheung, G., & Rensvold, R. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233-255.

Churchill, G. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *Journal Of Marketing Research*, 16(1), 64-73.

Cole, D., Ciesla, J., & Steiger, J. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, 12(4), 381-398.

Didellon, L., & Valette-Florence, P. (1996). L'utilisation des indices d'ajustement dans les modèles d'équations structurelles : présentation et recommandations d'usage. *Actes Des 12E Journées Nationales Des IAE*, Toulouse, 111-125.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 56-83.

Fornell, C., & Larcker, D. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal Of Marketing Research*, 18(1), 39-50.

Friborg, O., Martinussen, M., & Rosenvinge, J. (2006). Likert-based vs. semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience. *Personality And Individual Differences*, 40(5), 873-884.

Gefen, D., Straub, D., & Boudreau, M. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications Of The Association For Information Systems*, 4, Article 7.

Hair, J., Black, W., Babin, B., & Anderson, R. (2010). *Multivariate data analysis, a global perspective* (7th ed.). New Jersey: Pearson Prentice Hall.

Houston, M., & Rothschild, M. (1978). Conceptual and Methodological Perspectives on Involvement. In S.C. Jain, *1978 Educators' Proceedings*, (ed. pp. 186-187). Chicago: American Marketing Association.

Hu, L., & Bentler, P. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424-453.

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.

Jain, & Srinivasan. (1990). An Empirical Assessment of Multiple Operationalizations of Involvement. *Advances In Consumer Research*, 17, 594-602.

Jennrich, R. (2007). Rotation methods, algorithms, and standard errors. In R. Cudeck & R. MacCallum, *Factor analysis at 100: Historical developments and future directions* (1st ed., pp. 315-335). NJ: Erlbaum, Mahwah.

Kline, R. (2011). *Principles and Practice of Structural Equation Modeling* (3rd ed.). New York: Guilford Press.

Krugman, H. (1965). The Impact of Television Advertising: Learning Without Involvement. *Public Opinion Quarterly*, 29(3), 349-356.

Lastovicka, J., & Gardner, M. (1979). Components of Involvement. In C. Maloney & B. Silverman, *Attitude Research Plays for High Stakes* (pp. 53-73). Chicago: American Marketing Association.

Laurent, G., & Kapferer, J. (1985). Measuring Consumer Involvement Profiles. *Journal Of Marketing Research*, 22(1), 41-53.

Little, R. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal Of The American Statistical Association*, 83(404), 1198-1202.

Lynch, J. (1983). The Role of External Validity in Theoretical Research. *Journal of Consumer Research*, 10(1), 109-114.

MacCallum, R., & Austin, J. (2000). Applications of Structural Equation Modeling in Psychological Research. *Annual Review Of Psychology,* 51(1), 201-226.

Malhotra, N., Decaudin, J., Bouguerra, A., & Bories, D. (2011). *Études marketing* (1st ed.). Paris: Pearson.

Marsh, H., Liem, G., Martin, A., Morin, A., & Nagengast, B. (2011). Methodological Measurement Fruitfulness of Exploratory Structural Equation Modeling (ESEM): New Approaches to Key Substantive Issues in Motivation and Engagement. *Journal Of Psychoeducational Assessment*, 29(4), 322-346.

Marsh, H., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, 22(3), 471-491.

Marsh, H., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A., & Trautwein, U. (2009). Exploratory Structural Equation Modeling, Integrating CFA and EFA: Application to Students' Evaluations of University Teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 439-476.

McQuarrie, E., & Munson, M. (1992). A revised product involvement inventory improve usability and validity. *Advances In Consumer Research*, 19, 108-115.

Mindak, W. (1961). Fitting the Semantic Differential to the Marketing Problem. *Journal Of Marketing*, 25(4), 28-33.

Mitchell, A. (1979). Involvement: A Potentially Important Mediator of Consumer Behavior. *Advances In Consumer Behavior*, 6, 191-196.

Mittal, B. (1989). A Theoretical Analysis of Two Recent Measures of Involvement. *Advances In Consumer Research*, 16, 697-702.

Mittal, B. (1995). A comparative analysis of four scales of consumer involvement. *Psychology And Marketing*, 12(7), 663-682.

Mulaik, S. (1987). A Brief History of the Philosophical Foundations of Exploratory Factor Analysis. *Multivariate Behavioral Research*, 22(3), 267-305.

Perry, J., Nicholls, A., Clough, P., & Crust, L. (2015). Assessing Model Fit: Caveats and Recommendations for Confirmatory Factor Analysis and Exploratory Structural Equation Modeling. *Measurement In Physical Education And Exercise Science*, 19(1), 12-21.

Raftery, A. (1993). Bayesian model selection in structural equation models. In K. Bollen & J. Long, *Testing structural equation models* (1st ed., pp. 163-180). Newbury Park, CA: Sage.

Rosellini, A., & Brown, T. (2010). The NEO Five-Factor Inventory: Latent Structure and Relationships With Dimensions of Anxiety and Depressive Disorders in a Large Clinical Sample. *Assessment*, 18(1), 27-38.

Rossiter, J. (2002). The C-OAR-SE procedure for scale development in marketing. International *Journal Of Research In Marketing*, 19(4), 305-335.

Rothschild, M. (1984). Perspectives on Involvement: Current Problems and Future Directions. In T. Kinnear, NA - *Advances in Consumer Research* (pp. 216-217). Provo, Utah: Association for Consumer Research.

Sass, D., & Schmitt, T. (2010). A Comparative Investigation of Rotation Criteria Within Exploratory Factor Analysis. *Multivariate Behavioral Research*, 45(1), 73-103.

Schmitt, T., & Sass, D. (2011). Rotation Criteria and Hypothesis Testing for Exploratory Factor Analysis: Implications for Factor Pattern Loadings and Interfactor Correlations. *Educational And Psychological Measurement*, 71(1), 95-113.

Schumacker, R., & Lomax, R. (2010). *A beginner's guide to structural equation modeling*. New York: Routledge.

Stenling, A., Ivarsson, A., Hassmen, P., & Lindwall, M. (2015). Using bifactor exploratory structural equation modeling to examine global and specific factors in measures of sports coaches' interpersonal styles. *Frontiers In Psychology*, 6, 1303.

Zaichkowsky, J. (1985). Measuring the Involvement Construct. *Journal of Consumer Research*, 12(3), 341-352.

Zaichkowsky, J. (1986). Conceptualizing involvement. *Journal Of Advertising*, 15(2), 4-14.

Zaichkowsky, J. (1994). The Personal Involvement Inventory: Reduction, Revision, and Application to Advertising. *Journal Of Advertising*, 23(4), 59-70.

4.7 APPENDICES

Appendix 5. Information on the survey (faculty, place, classroom, course, observations, date)

Faculty of Literature and Human Sciences	**Faculty of Law**	**Faculty of Economics and Business**	**Faculty of Science**
(Espace Louis-Agassiz 1, CH-2000 Neuchâtel)	(Av. du 1er-Mars 26, CH-2000 Neuchâtel)	(Rue A.-L. Breguet 2, CH-2000 Neuchâtel)	(Rue Emile-Argand 11, CH-2000 Neuchâtel)
Place: Grand auditoire Class: Bachelor Course: Histoire de l'ethnologie Teacher: Dr. Nolwenn Bühler Number of observations: 17 Date: 26/04/2016	Place: B29 Class: Master Course: Biotechnologies et environnement juridique des sciences de la vie Teacher: Prof. Daniel Kraus Number of observations: 15 Date: 29/04/2016	Place: D67 Class: Bachelor Course: Management Teacher: Prof. Sam Blili Number of observations: 46 Date: 04/05/2016	Place: F200 Auditoire Louis Guillaume Class: Bachelor Course: Biologie des insectes Teacher: Dr. Praz Christophe Number of observations: 36 Date: 26/04/2016
Place: Aula Jeune Rives Class: Bachelor Course: Sciences de l'information et de la communication Teacher: Dr. Laurent Moccozet Number of observations: 15 Date: 27/04/2016	Place: C46 Class: Bachelor Course: Droit des obligations Teacher: Prof. Blaise Carron Number of observations: 52 Date : 11/05/2016	Place: R 107 Class: Master Course: Déontologie et éthique de l'information Teacher: Prof. Benoit Grevisse Number of observations: 25 Date: 12/05/2016	Place: F-100 Aula Unimail Class: Bachelor Course: Physiologie générale Teacher: Prof. Anne Prévot Number of observations: 121 Date: 03/05/2016
Place: Aula Jeune Rives	Place: C46 Class: Bachelor		

Class: Bachelor Course: Linguistique général Teacher: Dr. Christina Grisot Number of observations: 90 Date: 12/05/2016	Course: Droit pénal général Teacher: Prof. Kuhn André Number of observations: 65 Date: 03/05/2016		
Place: R.N.02 Class: Bachelor Course: Introduction à la sociologie économique Teacher: Prof. Olivier Crevoisier Number of observations: 30 Date: 27/05/2016			

Appendix 6. Survey questionnaire (French translation)

Questionnaire sur l'implication des consommateurs pour les chaussures

Utilisation du questionnaire

Dans le cadre d'une étude doctorale, ce questionnaire, ***sans aucun but lucratif****, a pour objet d'analyser le comportement du consommateur pour un produit particulier, les chaussures. Les réponses que vous donnez* ***sont strictement confidentielles****. Les données recueillies ne feront l'objet que d'un traitement statistique global. Il n'y a pas de réponse juste ou fausse; il faut seulement* ***répondre en toute franchise****. Répondez s'il vous plaît à toutes les questions. L'enquête prendra environ 10 minutes.*

Eric Rakotoasimbola

Question 1 : *On vous demande votre opinion sur un certain nombre de caractéristiques de votre situation d'achat d'une paire de chaussures. Sur une échelle de 1 à 5, veuillez SVP cochez pour chaque question la réponse correspondant le plus à votre opinion. Une seule réponse par échelle, comme suit :*

Vos réponses 1.*Pour moi, choisir une paire de chaussures est :*

R1. 1	Important	1	2	3	4	5	Pas important
R1. 2	Ennuyeux	1	2	3	4	5	Intéressant
R1. 3	Pertinent	1	2	3	4	5	Pas pertinent
R1. 4	Impliquant	1	2	3	4	5	Pas impliquant
R1. 5	Une activité excitante	1	2	3	4	5	Une activité pas excitante
R1. 6	Une activité géniale	1	2	3	4	5	Une activité pas géniale
R1. 7	Amusant	1	2	3	4	5	Pas amusant
R1. 8	Un fait marquant	1	2	3	4	5	Un fait peu marquant

Vos réponses 2. Pour moi, à mon avis, choix d'une paire de chaussure :

R2. 1	A une grande signification	1	2	3	4	5	N'a aucune signification
R2. 2	Me valorise	1	2	3	4	5	Ne me valorise pas
R2. 3	M'importe peu	1	2	3	4	5	M'importe énormément

Vos réponses 3. Pour moi, à mon avis, les chaussures :

R3. 1	M'attirent	1	2	3	4	5	Ne m'attirent pas
R3. 2	Sont des produits banals	1	2	3	4	5	Ne sont pas des produits banals
R3. 3	Sont des produits dont je n'ai pas besoin	1	2	3	4	5	Sont des produits dont j'ai besoin

Question 2 : *On vous cite différentes affirmations au sujet de l'achat de chaussures. Pour chacune d'entre elles, veuillez me dire si vous êtes pas du tout d'accord (note 1), plutôt en désaccord (note 2), ni d'accord ni en désaccord (note 3), plutôt d'accord (note 4), tout à fait d'accord (note 5).*

	Pas du tout d'accord	***Plutôt en désaccord***	***Ni d'accord ni en désaccord***	***Plutôt d'accord***	***Tout à fait d'accord***
1- Quand on achète des chaussures, on n'est jamais certain de son choix	1	2	3	4	5
2- Les chaussures que j'achète reflètent le genre de personne que je suis	1	2	3	4	5
3- On peut dire que les sujets sur les chaussures m'intéressent	1	2	3	4	5
4- Si après avoir acheté des chaussures, mon choix se révélait mauvais cela m'ennuierait énormément	1	2	3	4	5
5- J'attache énormément d'importance aux chaussures	1	2	3	4	5

6- Quand on choisit des chaussures, cela n'est pas grave si on se trompe	1	2	3	4	5
7- C'est très ennuyeux d'acheter des chaussures qui ne conviennent pas	1	2	3	4	5
8- Le choix d'une chaussure dit un peu qui on est	1	2	3	4	5
9- Pour moi, une chaussure c'est un peu un plaisir	1	2	3	4	5
10- Les chaussures, c'est un sujet qui me laisse totalement indifférent	1	2	3	4	5
11- On peut se faire une idée de quelqu'un aux chaussures qu'il a choisit	1	2	3	4	5
12- Quand je suis devant un rayon de chaussure, je me sens toujours désorienté(e) pour choisir	1	2	3	4	5
13- Je me fais un plaisir en achetant des chaussures	1	2	3	4	5
14- Choisir des chaussures c'est assez compliqué	1	2	3	4	5
15- Quand on achète des chaussures, on ne sait jamais si c'est bien celui-là qu'il fallait acheter	1	2	3	4	5
16- Quand on achète une chaussure on se fait un cadeau	1	2	3	4	5

Question 3 : *Sur une échelle de 1 à 5, veuillez cocher pour chaque énoncé votre opinion au sujet des énoncés suivants :*

Enoncé 1. *Généralement, je me renseigne régulièrement sur les nouvelles tendances de la mode*

Jamais	*Rarement*	*Quelquefois*	*Assez souvent*	*Presque toujours*
1	*2*	*3*	*4*	*5*

Enoncé 2. *J'aime me renseigner sur les chaussures sur internet / magazines / autres moyens d'informations*

Pas du tout d'accord	*Plutôt en désaccord*	*Ni d'accord ni en désaccord*	*Plutôt d'accord*	*Tout à fait d'accord*
1	*2*	*3*	*4*	*5*

Enoncé 3. *Je prête une attention particulière aux publicités des chaussures*

Pas du tout d'accord	*Plutôt en désaccord*	*Ni d'accord ni en désaccord*	*Plutôt d'accord*	*Tout à fait d'accord*
1	*2*	*3*	*4*	*5*

Enoncé 4. *Je passe beaucoup de temps pour choisir une paire de chaussures*

Pas du tout d'accord	*Plutôt en désaccord*	*Ni d'accord ni en désaccord*	*Plutôt d'accord*	*Tout à fait d'accord*
1	*2*	*3*	*4*	*5*

Question 4 *: Sélectionnez pour chaque énoncé la réponse qui vous correspond le plus*

Enoncé 1. *En général, j'achète des paires de chaussures :*

- ☐ Par mois
- ☐ Début de chaque saison
- ☐ A la période des soldes / Fin de saison
- ☐ Par année
- ☐ Quand j'en ai besoin
- ☐ Quand je trouve un modèle qui me convient
- ☐ Autres

Enoncé 2. *Chaque année pour des chaussures, je dépense environ :*

☐ 0 CHF - 250 CHF

☐ 251 CHF - 500 CHF

☐ 501 CHF - 750 CHF

☐ 1000 CHF et plus

Enoncé 3. *De manière générale, j'achète mes chaussures :*

☐ Aux magasins

☐ Aux supermarchés/hypermarchés

☐ Sur internet

Question 5 : *Veuillez classifier de* ***1 (le plus important)*** *à* ***7 (le moins important)*** *les critères suivants selon le degré d'importance que vous accordez pour déterminer le choix d'une paire de chaussure. Veillez à ne pas faire d'ex aequo.*

	Confortable	Fashion/A la mode	Marque prestigieuse	Esthétique	Durable	Economique	De Qualité
Classement							

Question 6 : *On souhaite connaître certaines informations sociodémographiques de base à votre sujet, nous vous remercions de nous indiquer* ***:***

1. Votre Age

☐

2. Votre état civil :

☐ Célibataire

☐ En couple

☐ Divorcé(e)

☐ Veuf/Veuve

3. Votre sexe :

- [ ] Femme
- [ ] Homme

4. De quelle faculté êtes-vous ?

- [ ] Faculté de droit
- [ ] Faculté des lettres et sciences humaines
- [ ] Faculté des sciences
- [ ] Faculté des sciences économiques

MERCI POUR VOTRE IMPORTANTE CONTRIBUTION A NOTRE RECHERCHE !

Appendix 7. Univariate statistics for each measurement model's items.

	N	Mean	Std.deviation	Missing		Number of extremes[a]	
				Count	Percent	Low	High
x1	509	2,24	1,103	3	,6	0	0
x2	509	2,64	1,206	3	,6	0	48
x3	502	2,75	1,084	10	2,0	0	35
x4	510	2,94	1,096	2	,4	0	0
x5	508	2,75	1,066	4	,8	0	37
x6	506	3,25	1,020	6	1,2	33	0
x7	506	3,66	1,165	6	1,2	0	0
x8	508	3,42	1,151	4	,8	40	0
x9	508	3,04	1,270	4	,8	0	0
x10	510	2,51	1,182	2	,4	0	39
x11	505	3,27	1,163	7	1,4	0	0
x12	505	4,34	,923	7	1,4	29	0
x13	503	2,71	1,055	9	1,8	0	37
x14	508	2,89	1,156	4	,8	0	0
x15	507	3,18	1,216	5	1,0	0	0
y1	511	3,88	1,000	1	,2	63	0
y2	509	2,26	1,046	3	,6	0	15
y3	510	4,11	1,018	2	,4	46	0
y4	509	2,53	1,179	3	,6	0	0
y5	508	2,82	1,093	4	,8	0	0
y6	509	2,64	1,206	3	,6	0	48
y7	510	2,67	1,161	2	,4	0	0
y8	509	2,59	1,122	3	,6	0	26
y9	509	3,26	1,110	3	,6	0	0
y10	510	2,71	1,067	2	,4	0	0
y11	510	3,49	1,043	2	,4	30	0
y12	510	3,62	1,023	2	,4	21	0
y13	509	3,13	1,186	3	,6	0	0
y14	509	3,40	1,136	3	,6	35	0
y15	509	3,62	1,012	3	,6	27	0
y16	510	3,75	1,088	2	,4	30	0

a. Number of cases outside the range (Q1 - 1,5*IQR, Q3 + 1,5*IQR).

Appendix 8. Results for comparisons between nomological models, sorted by method and by structural dimension

Nomological model comparison		$\Delta\chi^{2*}$	Δdf	$\chi^{2*}/\Delta df$	*-2*ΔLog MLR*	*-2*ΔLog MLR/Δdf*	*ΔAIC*	*ΔBIC*	*ΔRMSEA*	*ΔCFI*	*ΔTLI*	*ΔSRMR*
CIP	*Full model: CIP with ESEM* *Against* *Nested model : CIP with CFA*	100,763	44	2,290	100,886	2,292	-42,501	143,986	-0,004	0,018	0,008	-0,017
PII	*Full model: PII (2 dimensions) with ESEM* *Against* *Nested model : PII (1 dimension) with ESEM*	31,668	10	3,166	31,651	3,165	-37,623	-3,717	0,017	-0,028	-0,053	-0,012
	Full model: PII (2 dimensions) with CFA *Against* *Nested model : PII (1 dimension) with CFA*	43,357	2	21,678	43,477	21,738	-52,868	-44,391	-0,009	0,024	0,025	-0,003
	Full model: PII (2 dimensions) with ESEM *Against* *Nested model: PII (2 dimensions) with CFA*	13,332	8	1,666	13,284	1,660	-37,623	-3,717	0,017	-0,028	-0,053	-0,012
Modified PII	*CFA and ESEM are similar*	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
RRPII	*Full model: RRPII (2 dimensions) with ESEM* *Against* *Nested model : RRPII (2 dimensions) with CFA*	89,238	8	11,154	89,385	11,173	-112,558	-78,651	-0,028	0,039	0,047	-0,025

CHAPTER V. CONCLUSION

This study's primary goal is to provide specific answers to the following research questions:

- **ESSAY I:** How can we improve SEM recommendations and practices in consumer behavior research in Marketing?
- **ESSAY II:** How do conclusions based on fit indices vary relative to sample size, estimation method, and degree of nonnormality within the A-I-C model of Mittal and Lee (1989)?
- **ESSAY III**: For assessing the psychometric quality of consumer involvement measurement models, are there different results between Confirmatory Factor Analysis (CFA) and Exploratory Structural Equation Modeling (ESEM)?

The contributions and the limitations of the research are presented in the following sections.

5.1 CONTRIBUTIONS OF THE RESEARCH

5.1.1 Methodological contributions

The first essay has an educational purpose and is aimed at applying SEM rigorously. An eight-step decision protocol is suggested. Each step is a synthesis of recommendations and good practices for using SEM. The decision protocol is simple and intuitive, unlike a tailored case-by-case approach which requires the novice researcher to order the steps himself or herself, and which might be confusing. It also presents more coherence in the steps compared to other reviews of SEM in the marketing discipline. This study not only explains what to do but also shows how to do it. Finally, the decision protocol can be used as a checklist for assessing SEM applications for novice researchers, experts, journal editors, and practitioners.

The second essay aims at assessing the impact of sample size, degree of nonnormality, and estimation method on fit indices. Monte Carlo simulation was used to assess this impact. Results show the extent to which fit indices can diverge from each other when used to assess a model's fit to the data. New recommendations on the choice of fit indices for assessing A-I-C models were provided in this study. Root mean square error of approximation (RMSEA) and Standardized root mean square residual (SRMR) are recommended to assess the model, while Comparative fit index (CFI) and Tucker-Lewis index (TLI) required respectively at least 150 and 300 observations to be applied when assessing the model. Adjusted goodness-of-fit statistic (AGFI) and Goodness-of-fit statistic (GFI) are not recommended. The estimation method rarely affects fit indices when using any of the maximum likelihood methods. Using the generalized least squares (GLS) estimation method is not recommended as it overestimates the results for the TLI, SRMR, and CFI measures. The degree of nonnormality has a non-significant effect on the fit index, except for chi-squared, which is the measure it affects most. This example of applying the Monte Carlo simulation method shows how we can assess the suitability of fit indices by sample size, estimation, and other major traits likely to affect the fit indices.

The third essay illustrates the comparison of two structural equation approaches, the ESEM and the CFA, across four consumer product involvement models, namely consumer involvement profiles (CIP), product involvement inventory to advertising (PIIA), revised - revised product involvement inventory (RRPII), and modified PII. Results show that CFA and ESEM diverge in confirming discriminant validity for PIIA and CIP. This poses the question of which methods is reliable. The use of ESEM allows us to suggest items for CPI, RRPII, and PIIA that should be reconsidered or dropped from the model. The translation must be re-checked to determine what significance the consumer attaches to these items. Particularly for RRPII, cross-loading identification helps researchers identify possible improvement to achieve discriminant validity. By removing problematic cross-loadings, the two-dimensional RRPII shows discriminant validity.

Finally, we cannot conclude that ESEM can replace CFA, since ESEM encounters certain problems such as the calculation of the average variance extracted (AVE), which is one of the conventional techniques to assess convergent validity. To the best of our knowledge, there is no current technique to obtain equivalent data for ESEM. It was therefore necessary to take the CFA results as a reference.

5.1.2 Theoretical contributions

Most of this study is devoted to methodological issues in the use of SEM, but some theoretical contributions are obtained in Essay 3. Essay 3 helps us to obtain a ranking of four well-known consumer involvement measures. This ranking will help researchers select measurement models based on their performance on several psychometric criteria: reliability, convergent validity, discriminant validity, and nomological validity. The results of this study show that the PIIA was rejected by our data. This questioned its validity to be a generic measurement model of consumer involvement. The CIP model and RRPII model proved to be better than the other two measures, which are PIIA and modified PII.

5.2 LIMITATIONS OF THE RESEARCH

5.2.1 Methodological limitations

The first essay is devoted to standard SEM methods applied to cross-sectional data. Other advanced methods, such as longitudinal and experimental approaches to SEM, are not considered. We based our critique on the use of SEM in four research papers: Baumgartner & Homburg (1996); Hulland, Chow & Lam (1996); Martínez-López, Gázquez-Abad, & Sousa (2013); Chin, Peterson & Brown (2008); and Richter, Sinkovics, Ringle, & Schlägel (2016). The three first papers based their longitudinal studies of practice of SEM on four high ranked marketing journals: Journal of Marketing Research (JMR), Journal of Consumer Research (JCR), Journal of Marketing (JM), and International Journal Research of Marketing (IJRM) (Martínez-López et al., 2013; Baumgartner and Homburg, 1996). A much broader base would have been desirable. Finally, a synthesis of recommendations regarding SEM from other disciplines were outside the essay's scope.

In the second essay, the results of the Monte Carlo simulation method depend on the independent variables considered, but also on the characteristics of the model being tested, such as the number of relations to be tested. Therefore, the results of this study cannot be used as a rule of thumb for other models. The results are only relevant for a model which is similar to the A-I-C model in terms of the structural relationship and number of variables. We would have liked to work with the original data used by researchers applying SEM, so as to better replicate their methods in applying SEM and to test different techniques which might affect the results of the fit indices. Unfortunately, we received no satisfactory replies to the 150 e-mail inquiries we sent. Unlike researchers in other fields such as biology, researchers studying consumer behavior research in marketing -- at least the ones we contacted-- proved reluctant to share their data. The main cited reasons for refusing to share data are the loss of the files, confidentiality, or work in progress. The only dataset we managed to obtain was useless because the file was damaged.

In Essay 3, there are two main methodological limitations. The first relates to the choice of measurement models, since no single measurement model can be considered representative of all measurement models. Indeed, there is no single model for which statistical tests conducted on it can be generalized to all other models. The second limitation is that this study focuses only on comparing the results from two methods (CFA and ESEM). Further studies to assess the performance of each method's population values are needed. In this regard, Monte Carlo simulations can be used.

5.2.2 Theoretical limitations

As previously mentioned, this study focuses on the methodology for applying SEM. The interpretations yielded by the conceptual models tested in this study are mainly restricted to issues of statistical interpretation. Explanations of statistical results that diverge from the initial models must be interpreted by relying on theory. Such interpretations would exceed the scope of this study, since we chose the concept of involvement only for the limited purpose of illustrating a methodology.

5.3 SUGGESTIONS FOR FUTURE RESEARCH

To extend the first essay, we plan to do a Delphi survey with a panel of SEM experts and practitioners to assess and to improve the reliability of the decision protocol. Then, the decision protocol can be implemented in a statistical software for researchers and practitioners in the data science industry. Finally, an expanded version of the decision protocol for advanced SEM techniques (e.g. longitudinal SEM, nonlinear SEM) could be also developed.

The second essay is devoted to assessing fit indices performance based on Mittal and Lee's (1989) A-I-C model and Monte Carlo simulation. Recommendations for the selection of fit indices and some practical ways to assess the model with empirical data are provided. As it is known that the model was statistically rejected by fit indices in some studies, these studies should be re-assessed in regards to the new proposed recommendations.

For the third essay, first we plan to divide this study into two parts, then develop two papers: one paper for theoretical contributions and one paper for methodological contributions. One part will assess the four consumer involvement models based on the collected data using only ESEM. This part will highlight the empirical contribution of ranking the four consumer involvement measurement models. In addition, it will include more extended discussion of the models.

Alternatively, replications of each of the four models with ESEM can also be done, each of which could be a separate paper. The other part will only focus on the comparison of CFA and ESEM based on the same data. This will eventually produce a paper which only focuses on methodological issues. Finally, a future study can evaluate the performance of the Monte Carlo simulation to compare the two methods. In this regard the four involvement models can again be used as models to be studied.

5.4 REFERENCES

Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal Of Research In Marketing*, *13*(2), 139-161.

Chin, W., Peterson, R., & Brown, S. (2008). Structural Equation Modeling in Marketing: Some Practical Reminders. *The Journal Of Marketing Theory And Practice*, *16*(4), 287-298.

Hulland, J., Chow, Y., & Lam, S. (1996). Use of causal models in marketing research: A review. *International Journal Of Research In Marketing*, *13*(2), 181-197.

Martínez- López, F., Gázquez- Abad, J., & Sousa, C. (2013). Structural equation modeling in marketing and business research. *European Journal Of Marketing*, *47*(1/2), 115-152.

Mittal, B., & Lee, M. (1989). A causal model of consumer involvement. *Journal Of Economic Psychology*, *10*(3), 363-389.

Richter, N., Sinkovics, R., Ringle, C., & Schlägel, C. (2016). A critical look at the use of SEM in international business research. *International Marketing Review*, 33(3), 376-404.