Journal of Learning Theory and Methodology
Журнал теорії та методології навчання

ISSN 2708-7581 (Online)
ISSN-L 2708-7573

Methodological Review and Comparative Analysis

Methods for Estimating Sample Size in Structural Equation Modeling

Fernando Almeida[1ABCD]

[1]University of Porto & INESC TEC

Authors' Contribution: A – Study design; B – Data collection; C – Statistical analysis; D – Manuscript Preparation; E – Funds Collection

DOI: 10.17309/jltm.2026.7.2.03

Abstract

Background. Sample size estimation in Structural Equation Modeling (SEM) remains a debated methodological issue. Although minimum sample size rules are widely used, they often lack integration of theoretical, statistical, and practical considerations, particularly in complex research designs.

Objectives. This study aims to provide a synthesis and comparative analysis of different approaches to sample size estimation in SEM, integrating theoretical foundations, statistical criteria, methodological limitations, and applied research contexts.

Materials and Methods. A multi-method approach was employed, consisting of three components: a narrative review of sample size estimation methods, consideration of simulated scenarios reflecting realistic research conditions, and a comparative evaluation of method performance across these scenarios.

Results. Findings show that rule-of-thumb approaches are useful for preliminary estimates but may misrepresent required sample sizes in complex models. Power analysis provides more precise and rigorous estimates, particularly for confirmatory research. Monte Carlo simulations offer robust guidance for complex, longitudinal, and multigroup models. Estimator-based recommendations highlight the importance of aligning sample size with data characteristics and estimation techniques. Model complexity emerged as a key determinant, with advanced SEM models often requiring 400–700+ participants.

Conclusions. A systematic and context-sensitive approach to sample size estimation is essential for ensuring robustness and methodological rigor in SEM studies across disciplines such as management and education.

Keywords: research methodology, structural equation modeling, sample size estimation, statistical power, model complexity.

Introduction

Structural Equation Modeling (SEM) is an advanced statistical approach that has gained enormous relevance in various fields such as engineering, management, and social sciences, mainly due to its ability to deal with complex relationships between observable and latent variables. Nusair and Hua (2010) and Xiong et al. (2015) confirm this view by pointing out that its importance stems from the fact that it allows researchers to test theories robustly, integrating multiple constructs that cannot be measured directly (e.g., attitudes, perceptions, trust, intentions, or skills). Thus, SEM offers a framework that combines confirmatory factor analysis and multiple regression models. This makes it possible to simultaneously assess the quality of measures and the structural relationships between constructs.

The applicability of SEM is particularly evident when researchers deal with multidimensional theoretical models, in which simple linear regression is insufficient. For example, in studies in the field of innovation, entrepreneurship, and organizational management, phenomena often involve interactions between psychological, behavioral, and contextual factors (Allammari et al., 2025; Dulaimi et al., 2005; Ghardashi et al., 2022). Its application allows for testing complex hypotheses, such as direct, indirect, and mediating effects, as well as moderating effects, providing a more holistic view of the processes studied. Furthermore, it makes it possible to compare alternative models and assess which one best fits the data, contributing to the validation of emerging theories. Another relevant aspect is the ability to work with latent variables, which reduce measurement error by aggregating multiple indicators to represent a construct (Kline, 2010). This feature provides greater accuracy to the results and improves the internal validity of the models. SEM also makes it possible to assess the quality of the scales used through metrics such as composite reliability, convergent and discriminant validity, ensuring that the model measures what

JLTM
LLC OVS

Corresponding Author: Fernando Almeida,
e-mail: almd@fe.up.pt

it is intended to measure (Cheung et al., 2024; Newsom, 2023; Panzeri et al., 2024).

Determining the minimum sample size for applying SEM is one of the most critical aspects of ensuring the validity and robustness of the results. The importance of this definition stems from the fact that SEM is a complex statistical technique that involves multiple parameters to be estimated, including factor loadings, covariances, structural paths, and measurement errors (Kline, 2010). The greater the number of parameters, the greater the need for sufficient data to ensure that the estimates are stable and that the model can converge without problems. An inadequate sample size can result in biased estimates, making it impossible to identify the model or leading to weak fit indices that do not adequately reflect the quality of the theory being tested. Furthermore, as Kim (2005) points out, sample size directly influences the statistical power of the model. Small samples may fail to detect significant effects even when they exist, which undermines the ability to validate theoretical relationships. However, determining this minimum size is a complex and challenging task. One of the difficulties lies in the absence of a universal consensus. The diversity of estimators available in SEM adds another layer of complexity, as different methods require different sample sizes. The presence of non-normal variables, complex models, or structures with many indicators also increases the need for larger samples. Another challenge lies in the dependence on the specific context. Two models with the same number of parameters may require different sample sizes depending on the quality of the measurements and the magnitude of the expected effects. Thus, determining the minimum sample size in SEM is not a simple application of rules, but rather a process that requires methodological judgment. In this sense, this study seeks to synthesize the various approaches to sample calculation and explore practical perspectives on their application, considering the context in which they can be used. Although there are studies that discuss minimum sample requirements, such as Hox et al. (2010), Nicolaou & Masoner (2013), and Ramos-Vera (2022), few offer an integrated view that combines theoretical foundations, statistical criteria, methodological limitations, and the practical translation of these guidelines into real research scenarios. This gap provides an opportunity for this study to offer a relevant and distinctive contribution.

The rest of this article is organized as follows: first, a theoretical contextualization of different methods for estimating the minimum sample size in SEM is presented. Next, the methodology adopted in this study is presented, indicating its three phases. After that, the results are presented, and their relevance and practical implications are discussed. Finally, the main conclusions are summarized and suggestions for future work are provided.

Literature Review

Calculating the minimum sample size in SEM involves a diverse set of methods that reflect both statistical advances and practical needs in empirical research. There is no universally accepted criterion, which makes the topic particularly complex and relevant. Literature in this field, such as Wang and Rhemtulla (2021) and Wolf et al. (2013), suggests that the choice of method should incorporate various elements, such as the type of model, the number of constructs considered, the variability of the sample data, and the type of estimator used. It is recognized that these different approaches offer complementary perspectives on what constitutes a sufficiently robust sample to ensure valid results.

One of the most traditional methods is that of practical rules based on ratios between the number of participants and the number of parameters or indicators. Among the best known is the recommendation of at least 10 cases per parameter or indicator, although some authors propose more conservative ratios, such as 15 or 20 cases per variable (Ali Memon et al., 2020; Iacobucci, 2010; Kush et al., 2021). Although simple, this approach has been criticized for not considering the complexity of the model or the specific statistical properties of the data. Even so, it continues to be used as an initial reference, especially in exploratory studies or in preliminary planning stages.

Another relevant approach is statistical power analysis, which seeks to determine the sample size needed to detect effects of a certain magnitude with desirable levels of confidence and power. This method uses previously estimated or expected values, such as factor loadings, correlations, and structural effects, allowing for a more rigorous assessment of the model's ability to detect real relationships in the data (Buchberger et al., 2024; Flora et al., 2025). Power analysis is particularly useful in confirmatory studies, but requires prior knowledge of the parameters, which is not always available.

Monte Carlo simulations represent one of the most advanced and methodologically robust approaches. Traditionally, a Monte Carlo simulation is viewed as a computational method used to understand the behavior of a system or model by running a large number of random simulations (Graf & Korn, 2020; Harrison, 2010). It relies on repeated random sampling to estimate how a model performs under different conditions. It is particularly useful to assess uncertainty, variability, or risk in predictions (Schamberger, 2023). In the context of SEM, Monte Carlo simulations are used to: (i) test how sample size affects parameter estimates and model fit; (ii) examine the impact of non-normal data, missing values, or measurement error; and (iii) evaluate power and stability of complex models, including those with second-order factors, multigroup comparisons, or longitudinal designs. Serang (2020) notes that this method allows for the consideration of complex factors, such as data non-normality and distinct factor loadings. Despite its high accuracy, the use of simulations requires more advanced statistical skills and specific tools, which limit its applicability.

An additional approach involves recommendations based on the type of estimator used. Estimators such as Maximum Likelihood, as discussed by Maydeu-Olivares (2017), tend to require larger samples, especially in non-normal contexts, while robust or asymptotic distribution-based methods can tolerate smaller samples, albeit with potential trade-offs in accuracy. The Weighted Least Squares Mean and Variance Adjusted (WLSMV) estimator, for example, is often recommended for categorical data, but requires relatively large sample sizes to ensure stability (DiStefano & Morgan, 2014).

Another perspective is based on the structural complexity of the model. Models with many constructs, multiple paths, or second-order latent variables require significantly larger

samples than simple models. Each of these parameters requires sufficient information to be calculated accurately. When the sample is small relative to the number of parameters, the model tends to exhibit instability, convergence difficulties, and unreliable estimates. In addition, fit indices such as CFI, TLI, or RMSEA become more sensitive to fluctuations, producing results that may mistakenly suggest a poor fit. van Kesteren & Oberski (2019) point out that complexity also increases when there are multiple structural paths between constructs. In models that test direct, indirect, moderating, and mediating effects simultaneously, the amount of information needed to capture these relationships grows exponentially. Indirect effects, for example, are particularly sensitive to sample size because they involve the multiplication of coefficients, which amplifies estimation errors. Thus, it can be concluded that small samples can lead to underestimation or overestimation of these effects, compromising theoretical conclusions. In addition, another factor of complexity arises that is related to second-order latent variables (Koufteros et al., 2009; Yang et al., 2020). These structures, which aggregate several first-order constructs, add new layers to the model, requiring more parameters and increasing the difficulty of stabilizing the model with few observations.

Materials and Methods

This study applies an exploratory methodology that aims to compare different methods for calculating the minimum sample size in SEM. This approach is particularly appropriate in a field where there is no consensus and where recommendations vary substantially depending on the model and research objectives. Thus, the proposed methodology recognizes that calculating sample size is not a purely technical issue, but rather a process that requires coordination between theory, practice, and empirical evaluation.

Figure 1 presents visually the methodological process. The first component of this approach involves a narrative literature review of the main existing methods. This step allows us to map different families of methods, such as practical rules, power analyses, criteria based on the number of parameters, and Monte Carlo simulations. By critically exploring their premises and contexts of applicability, the study provides a solid theoretical basis that helps to understand why different methods produce different recommendations. This review also clarifies gaps and inconsistencies in the literature, justifying the need for additional empirical comparisons.

The second component of the study is centered on the practical application of sample size determination methodologies through the development of structured simulation-based scenarios that reflect realistic research conditions. These scenarios are designed to mirror common challenges encountered in empirical studies, including varying levels of structural complexity such as first- and second-order latent variable models, the presence of mediation effects, multigroup comparisons, non-normal data distributions, and models incorporating categorical or ordinal indicators. However, it is important to clarify that, although the study adopts a simulation-oriented design logic, it does not present full Monte Carlo simulation outputs or empirically generated sampling distributions. Instead, the term "simulation" is used in a conceptual and scenario-building sense, where hypothetical but realistic model configurations are constructed to examine how different sample size estimation approaches behave under diverse methodological conditions. This includes exploring how practical rules, statistical power analysis, Monte Carlo-based reasoning, estimator-specific recommendations, and model complexity considerations may vary when applied to increasingly demanding analytical contexts. Therefore, each scenario functions as a controlled analytical framework rather than a computational simulation experiment, allowing for structured comparison across methodological approaches. Additionally, this study creates a set of analytically grounded situations that reflect the heterogeneity of contemporary empirical research in social sciences, management, education, and engineering.

The third component of the study consists of a structured comparative evaluation of the sample size estimates produced by the different methodological approaches across the defined scenarios. This analytical step goes beyond simple observation of convergence or divergence between methods, aiming instead to examine the magnitude, direction, and consistency of differences under varying model conditions. In practice, the comparison was conducted by applying each estimation approach (e.g., practical rules, statistical power analysis, estimator-based recommendations, and model complexity guidelines) to the same set of scenario specifications. This ensured a common analytical baseline, allowing results to be directly contrasted across methods within identical structural conditions. To enhance analytical rigor and reduce potential bias in interpretation, several strategies were adopted. First, all scenarios were defined a priori and kept constant across methods, preventing post hoc adjustments that could favor a particular approach. Second, comparisons were conducted using a standardized parameter framework, ensuring that differences in sample size recommendations were not driven by inconsistencies in model specification (e.g., number of latent variables, indicators, or structural paths). Third, the analysis emphasized relative comparison patterns rather than absolute numerical differences alone, focusing on whether methods systematically produced more conservative (larger sample requirements) or more liberal (smaller sample requirements) estimates under identical conditions. This assessment is not limited to a numerical comparison of the recommended values; it also involves analyzing the accuracy of the model estimates, the stability of the parameters, the behavior of the goodness-of-fit indices, and the ability of each method to deal with problems reported by Kelava et al. (2008) and Tarka (2017), such as non-normality, multicollinearity, or models with high structural complexity. In addition, this step allows us to assess the sensitivity of the methods to the specific characteristics of the scenarios. For example, if you want to explore how rule-based methods behave when the model contains second-order variables, or how power analysis reacts when the expected effects are small or moderate.

The combination of these three components fully justifies a multi-method approach. As Hammond (2005) points out, a multi-method approach combines two or more methods of data collection, analysis, or interpretation within the same project, with the aim of obtaining a more complete understanding of the phenomenon under study. Thus, rather than relying exclusively on a single method (e.g., whether quantitative, qualitative, or experimental), a multi-

method study integrates different methodological strategies to explore a research question from multiple perspectives. Applied to the specific context of this study, it is recognized that no single method can capture all the factors that influence the adequacy of the sample in SEM. Therefore, this multi-method methodology not only enriches the analytical rigor of the study but also increases its usefulness for researchers who need well-founded recommendations that are adaptable to the specific context of their investigations.

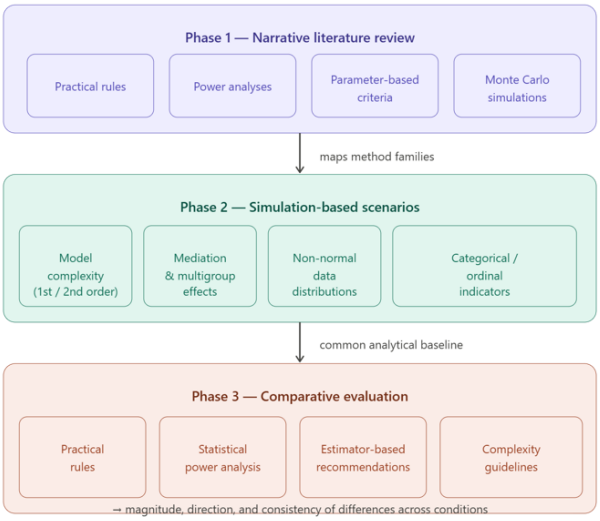

Fig. 1. Phases of the methodological process

Results

Table 1 provides a comparative analysis of different methods to sample size estimation in SEM. Findings reveal that each method for estimating sample size in SEM serves different purposes and fits different research contexts. Practical rules based on participant-to-parameter ratios function mainly as an initial approximation rather than a definitive guideline. They are best used at the earliest planning stage, offering a quick benchmark but lacking the precision required for rigorous studies. Statistical power analysis is more appropriate when the researcher has prior information about expected effect sizes and wishes to ensure sufficient power and confidence. This method is especially valuable in studies where the detection of specific relationships is crucial and where analytical rigor is required. Monte Carlo simulations are particularly suitable for advanced research and complex SEM models. Their strength lies in the ability to empirically test how the model behaves across different sample sizes and data conditions, making them ideal when precision and methodological robustness are priorities. Researchers dealing with innovative, intricate, or high-stakes models often rely on this approach. Recommendations based on the type of estimator used become relevant when the properties of the data are well defined. For instance, models estimated using Maximum Likelihood tend to perform adequately with moderate sample sizes under normality, making this approach suitable when such assumptions are met. Conversely, when data are categorical, the WLSMV estimator becomes the most appropriate choice, though it typically requires considerably larger samples to ensure stability and accuracy. Finally, considering the structural complexity of the model is essential when working with large-scale, multifactorial, or hierarchical SEM designs. As complexity increases, so does the required sample size, making this perspective indispensable for ensuring reliable estimation in sophisticated models.

Figure 2 provides a methodological flowchart that guides researchers in selecting the most appropriate sample-size estimation approach according to model characteristics, data properties, and research objectives. The diagram poses four sequential binary questions related to the researcher's

Table 1. Comparative Analysis of Sample Size Estimation Methods in SEM

Method	Description	Advantages	Limitations
Practical rules (participant-to-parameter ratios)	Based on fixed ratios, such as 10 to 20 participants per parameter or indicator.	Simple, quick, and widely used as an initial reference.	Ignores model complexity and effect sizes. May lead to under- or overestimated samples.
Statistical power analysis	Calculates the sample size needed to detect specific effects with a desired level of confidence and power.	More rigorous and tailored to expected effect sizes.	Requires precise definition of effect magnitudes.
Monte Carlo simulations	Tests model performance under different sample size, variability, and structural scenarios.	Highly robust and empirically grounded. Provides very precise estimates.	Requires software, computation time, and technical expertise. High level of complexity.
Estimator-based recommendations (e.g., ML)	Sample size depends on the characteristics and assumptions of the estimator used.	Can provide estimates better aligned with data characteristics.	Strongly dependent on assumption compliance. Different estimators vary widely in sample demands.
WLSMV for categorical data	Robust estimator for ordinal or categorical variables, with mean and variance adjustments.	Excellent for categorical data. Reduces bias associated with ML in non-continuous variables.	Requires large samples for stable estimates, especially with few categories or many factors.
Model structural complexity	Considers the number of constructs, paths, factor loadings, and presence of second-order latent variables.	Accounts for the actual structural demands of the model. Adjusts sample size to model complexity.	May require very large samples. Difficult to use in isolation without other methodologies.

situation and model characteristics. The first question determines whether the study is still in an early planning stage, directing those who are toward practical rules based on participant-to-parameter ratios. The second question asks whether expected effect sizes are known, routing researchers with that prior knowledge toward statistical power analysis. The third question identifies whether outcome variables are categorical or ordinal, pointing to the WLSMV estimator in such cases. The fourth and final question addresses whether the model is large-scale, multifactorial, or hierarchical, recommending a structural complexity approach when that is true. Researchers who answer no to all four questions are directed to the Maximum Likelihood estimator, suited for continuous data under normality assumptions. It is also recommended the combination of methods for greater precision reflecting the complementary nature of the approaches presented in Table 1.

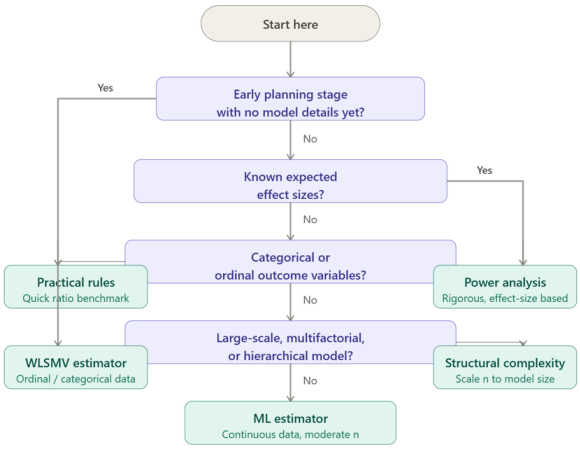

Fig. 2. Methodological flowchart for selecting sample size estimation approaches

The results are further explored considering practical scenarios for the use of SEM in areas such as engineering, management, social sciences, and education. It is important to demonstrate how SEM has broad applicability in multiple areas. Table 2 provides practical guidance on estimating sample sizes for SEM across different methods, scenarios, and fields. However, it is important to clarify that these recommendations are not derived from a single unified empirical dataset or formal simulation study, but rather from a structured synthesis of widely reported guidelines in the SEM literature, combined with established methodological conventions.

The synthesis indicates a relatively stable convergence of recommended minimum sample sizes across SEM traditions, with variability primarily driven by estimator choice and effect size assumptions. Across the reviewed methodological benchmarks, the practical rules approach (10–20 observations per estimated parameter) typically translates into approximate sample sizes ranging from 50–200 (simple models with 5–15 parameters) to 300–400 (more parameter-intensive models). This rule is consistent with early-stage exploratory SEM applications, where precision is secondary to model feasibility and identification.

The statistical power analysis framework, assuming a conventional medium effect size ($f^2 = 0.15$), $\alpha = 0.05$, and power = 0.80, produces more formalized estimates in the range of 150–250 participants for standard SEM models, increasing to 250–400 participants when structural complexity or lower effect sizes are considered. More specifically, this framework derives its robustness from its explicit grounding in Type I and Type II error control, ensuring that the probability of incorrectly rejecting a true model effect (α error) and failing to detect an existing effect (β error) is explicitly regulated. Compared to rule-of-thumb approaches, power analysis incorporates effect size sensitivity, meaning that even moderate reductions in assumed effect magnitude (e.g., from $f^2 = 0.15$ to $f^2 = 0.10$) can lead to substantial increases in required sample size, often exceeding 300–450 participants depending on model specification and number of latent constructs. In SEM contexts, these estimates are further influenced by model complexity parameters. Models with higher parameter density or weaker indicator loadings typically require larger samples to maintain sufficient statistical power for detecting both direct and indirect effects. In mediation models, for instance, indirect effects tend to have lower statistical power than direct effects, often necessitating an additional 20–40% increase in sample size to achieve equivalent detection capability. Moreover, power-based SEM recommendations assume adequate measurement quality, including acceptable reliability levels (e.g., factor loadings ≥ 0.60–0.70). When measurement error increases, the effective power of the model decreases, thereby inflating required sample sizes even further. This reinforces the importance of integrating measurement model assessment with power calculations, rather than treating sample size determination as a purely structural issue.

The Monte Carlo simulation literature (as reported in prior validated studies) indicates more robust and empirically stable requirements, typically ranging from 250–600 participants, with the upper bound associated with conditions of non-normality, higher model misspecification sensitivity, or longitudinal/multigroup structures. These estimates reflect convergence criteria such as parameter bias (<10%), standard error stability, and acceptable convergence rates (>95%), as commonly reported in simulation-based SEM validation studies. Effectively, Monte Carlo approaches derive their strength from the ability to replicate the sampling process under controlled population conditions. Within this framework, sample size requirements are not treated as fixed thresholds but as performance-dependent outcomes, defined by whether predefined statistical criteria are satisfied across a large number of replications (e.g., 1,000–5,000 simulated samples). In SEM applications, acceptable performance is typically evaluated using multiple diagnostics, including relative bias, root mean square error (RMSE), standard error accuracy, coverage probability of confidence intervals (typically ≥ 0.90–0.95), and empirical power stability across replications. When these criteria are jointly considered, sample size recommendations tend to increase as model complexity escalates. In particular, multigroup SEM models and longitudinal growth models tend to require larger samples (often 400–700+ observations) due to the need to simultaneously estimate group-specific or time-varying parameters. Similarly, models involving non-normal indicators or categorical data structures exhibit increased estimation instability at smaller sample sizes, often requiring upward adjustments in the recommended range to maintain convergence reliability and reduce parameter distortion.

From an estimator-specific perspective, Maximum Likelihood (ML) estimation demonstrates adequate performance with 200–300 observations under continuous and approximately normal data conditions, while the WLSMV estimator for ordinal/categorical indicators requires approximately 350–550 observations to maintain parameter stability and robust standard errors. These differences reflect documented sensitivity of categorical estimators to sample size and distributional asymmetry, particularly in models with multiple latent constructs. ML estimation is grounded in the assumption of multivariate normality and sufficient sample size for asymptotic properties to hold, meaning that parameter estimates become increasingly unbiased and efficient as sample size increases. Under conditions of mild to moderate non-normality, ML can still perform adequately, particularly when robust corrections (e.g., Satorra–Bentler scaled statistics or robust standard errors) are applied. However, when sample sizes approach the lower bound of the recommended range (≈200), ML estimates become more sensitive to model misspecification and measurement error, potentially affecting fit indices and standard error precision. In contrast, WLSMV (Weighted Least Squares Mean and Variance adjusted) estimation is specifically designed for ordinal or categorical indicators, where the underlying assumption is that observed responses reflect discretized versions of continuous latent response variables. This estimation approach relies heavily on accurate threshold estimation and polychoric correlations, both of which are highly sensitive to sample size. As a result, smaller samples can lead to unstable threshold estimates, inflated standard errors, and reduced convergence reliability, particularly in models with a large number of categories or indicators per latent construct.

Finally, model structural complexity introduces a consistent upward adjustment factor of approximately +30% to +60% in required sample size, particularly in second-order, multigroup, or longitudinal SEM designs. Under these conditions, recommended samples frequently reach 400–600 participants, with more complex longitudinal trajectories extending to 600–750 participants to ensure stable estimation of growth or group-specific parameters. This upward adjustment is primarily driven by the fact that structural complexity increases the total number of free parameters to be estimated, including additional factor loadings, structural paths, intercepts, variances, covariances, and, in multigroup settings, group-specific parameter sets. As the parameter-to-sample ratio becomes more constrained, estimation precision decreases unless compensated by larger samples. This effect is particularly pronounced in models with latent interactions, hierarchical factor structures, or cross-group equality constraints, where identification conditions require stronger empirical support. Moreover, in second-order SEM models, the introduction of higher-order latent constructs amplifies estimation demands because each higher-order factor aggregates multiple first-order constructs, thereby increasing the dependency structure within the model. This leads to greater sensitivity to sampling variability, especially when first-order factor loadings are moderate or unevenly distributed. Consequently, sample sizes closer to the upper bound (≈500–600) are typically required to maintain stable second-order parameter estimates and acceptable standard errors. Additionally, in multigroup SEM applications, sample size requirements are further influenced by the need to ensure measurement invariance across groups. Each additional group effectively reduces the available sample per estimated

Table 2. Practical Scenarios for the Application of SEM and Estimation the Sample Size

Method	Scenario I	Scenario II	Scenario III
Practical rules (participant-to-parameter ratios)	Exploratory research in startup incubators to map entrepreneurial competencies; ~150–200 participants (assuming 15 parameters).	Pilot with small software teams; ~80–120 participants for ~10 parameters.	Student satisfaction in a small class; ~50–100 participants for ~5–7 parameters.
Statistical power analysis	Leadership training effect on teams; ~200–250 participants to detect medium effect size (f^2 = 0.15) with 0.80 power.	Community program evaluation; ~180–220 participants to detect medium effects with $\alpha = 0.05$ and power 0.80.	Pedagogical intervention on student skills; ~250–400 participants to detect medium-to-large effects.
Monte Carlo simulations	Complex model that involves culture, networking, and innovation constructs; ~300–500 participants for stable parameter estimates.	Reliability of complex systems; ~250–400 participants depending on non-normality and latent variable structure.	Longitudinal urban mobility study; ~400–600 participants for stable growth trajectories.
Estimator-based recommendations (e.g., ML)	Customer satisfaction surveys with continuous data; ~300–400 participants for moderate complexity.	Industrial process comparison with continuous data from less than 3 sectors; ~200–300 participants.	Well-being factor analysis; ~350–500 participants for stable ML estimation.
WLSMV for categorical data	Likert-scale entrepreneurial competencies; ~400–500 participants due to ordinal items.	School climate evaluation; ~350–450 participants to ensure stability across multiple classes.	Hierarchical public policy perception models; ~450–550 participants for robust WLSMV estimation.
Model structural complexity	Second-order model of motivation, leadership, networking; ~400–600 participants.	Multigroup model comparing departments; ~250–400 participants depending on groups and paths.	Longitudinal model of student competencies; ~500–750 participants for trajectory estimation.

parameter, meaning that total sample size must scale with both the number of groups and the complexity of cross-group constraints.

Discussion

This study offers an analysis and comparison of different methods for estimating sample size in SEM. This is a central theme for ensuring validity, reliability, and methodological rigor in complex quantitative research. Several approaches are explored, ranging from simple practical rules to highly robust Monte Carlo simulations. The results obtained in this study are important for demonstrating how the choice of method directly impacts the accuracy of estimates, the stability of parameters, and statistical power. These are fundamental aspects for different fields of application, including management, engineering, education, and social sciences.

The study shows that traditional methods, such as practical rules based on the participant-parameter ratio, are useful in early stages or in exploratory research with limited resources, offering quick estimates of sample size, usually between 50 and 200 participants. However, current literature indicates that such approaches lack rigor and may underestimate or overestimate sample requirements in complex models (Schuberth et al., 2023). Thus, although useful, these estimates should be used sparingly and not abusively, particularly when the population size we are inferring is significant. Therefore, it can be concluded that the use of fixed rules can lead to underestimation or overestimation of sample size, especially in complex, multifactorial, or longitudinal models. Underestimating the sample can result in unstable parameters, high type I or type II errors, and unreliable conclusions (Akobeng, 2016; Darling, 2024; Rothman, 2010).

Statistical power analysis represents a significant methodological advance over traditional rules of estimation, as it allows one to calculate the sample size needed to detect effects of a specific magnitude with predetermined levels of confidence (usually $\alpha = 0.05$) and power (typically 0.80 or higher). Rather than providing only a rough estimate based on generic ratios, this approach explicitly considers the strength of the expected effects, which is critical to ensuring that statistical tests have a sufficient probability of identifying real relationships between variables, minimizing the risk of type II errors. This approach is particularly useful in confirmatory studies, in which the researcher already has clear hypotheses and wishes to verify specific relationships between constructs. For example, in corporate leadership programs, power analysis allows the calculation of the number of participants needed to detect whether a training intervention actually improves team performance or job satisfaction. Similarly, in educational interventions, such as critical skills teaching programs or active learning methods, power analysis helps define a sample capable of revealing statistically significant effects on academic performance or student motivation. Furthermore, power analysis is aligned with classic recommendations by McQuitty (2004), who emphasizes that studies with insufficient power tend to produce unreliable results, making it difficult to replicate and generalize conclusions. Subsequent SEM research, such as that by Cole (2025) and Irmer et al. (2024), reinforces that the use of power analysis is essential for confirmatory and multifactorial models, as it allows for balancing statistical rigor with practical feasibility, avoiding both undersizing (which compromises the detection of real effects) and oversizing (which can waste resources).

Monte Carlo simulations stand out as an empirically robust method capable of testing model stability in different scenarios of sample size and data distribution. More recent studies emphasize their value in complex or longitudinal models, especially when traditional assumptions of normality are not met (Hsu et al., 2019; Schultzberg & Muthén, 2017). Similarly, recommendations based on the type of estimator, such as Maximum Likelihood or WLSMV, reinforce that the adequacy of the sample size depends on the characteristics of the data. WLSMV, for example, requires considerably larger samples for categorical data, reflecting the need for alignment between model, data, and estimator.

Considering the structural complexity of the model is one of the most critical factors in defining the sample size in SEM, as more elaborate models place additional demands on the data and the accuracy of the estimates. Structural complexity refers to elements such as the number of latent constructs, the number of indicators per construct, the presence of second-order factors, multigroup structures, or longitudinal models that estimate trajectories over time. Each of these elements significantly increases the number of parameters to be estimated, which, in turn, increases the need for a larger sample to ensure statistical stability and reliability. For example, in second-order models, where higher-level latent constructs are formed from first-level factors, the number of paths and factor loadings to be estimated grows nonlinearly in relation to the number of observed variables. This implies that a sample that would be sufficient for a first-order model may be inadequate, resulting in unstable parameters, biased estimates, and wide confidence intervals. In multigroup models, complexity increases even further, as it is necessary to ensure stability in each group analyzed, in addition to allowing comparisons between groups, which require significantly larger samples. Longitudinal models, in turn, require sufficient participants to capture temporal trajectories, considering possible losses due to dropouts and variability over time. The current literature corroborates this need for larger samples in complex models. Studies such as those by Montoya & Edwards (2020) indicate that, for models with multiple factors, indicators, and groups, it is common for the required sample size to exceed 400 to 600 participants, reaching 700 or more in longitudinal or extensive second-order models. Moreover, Savalei (2020) shows that when using robust estimators or categorical data, this requirement may be even greater, as estimators need more information to produce reliable results.

Finally, we recognize the potential of this article in offering practical implications. First, it contributes to greater rigor in defining sample size. The aim is to help researchers plan data collection in a more scientific and informed manner, considering the characteristics of the model, structural complexity, type of data, and estimator used. This rigor avoids arbitrary decisions or those based solely on conventions, such as the generic rule of "10 to 20 participants per parameter." Although the sample size ranges presented

in this study are derived from established methodological literature and widely accepted SEM guidelines, they should be viewed as indicative reference values rather than universally applicable thresholds. Second, this study seeks to contribute to better resource allocation. SEM research can involve significant costs, both in terms of money and in terms of time and human effort. More accurate sample planning avoids unnecessarily large data collections, which consume resources without generating proportional gains in accuracy, and also avoids insufficient samples, which would require repeating studies or collecting additional data later. Thus, by adequately estimating the sample size, it is possible to optimize the budget, reduce logistical effort, and increase the operational efficiency of research projects. Furthermore, this situation is highly relevant in applied studies in management, engineering, education, or social sciences, where the availability of participants may be limited. Lastly, it also aims to contribute to ensuring robust and replicable results. An adequate sample size increases the stability of parameter estimates, reduces the risk of type I and type II errors, and improves the generalization of results to the target population. This is essential for confirmatory research and longitudinal studies, in which reliable conclusions depend on the consistency of estimates over time and across different groups. Furthermore, replicability, which has become a central criterion of scientific quality as demonstrated by Chakravorti et al. (2025), Epskamp (2019), and Rahal et al. (2022), is only possible when the sample allows real effects to be detected with statistical confidence. Although adequate sample size is a necessary condition for robust and replicable SEM results, it is not sufficient by itself. Replicability and statistical robustness also depend on factors such as model specification, measurement quality, construct reliability and validity, estimator appropriateness, treatment of missing data, and compliance with distributional assumptions (Guenther et al., 2025; Yuan et al., 2015). Consequently, the recommendations proposed in this study should be interpreted as methodological guidance that supports research planning rather than as guarantees of reproducible findings.

Conclusions

This study highlights the importance of adopting a systematic and careful approach to estimating sample size in SEM, showing that different methods offer varying levels of accuracy and applicability. Traditional rules of estimation are useful in exploratory phases, offering quick estimates, but they have limitations in complex models and can lead to underestimation or overestimation of sample requirements. Statistical power analysis represents an advance, allowing the calculation of samples capable of detecting specific effects with defined levels of confidence and power, and is particularly suitable for confirmatory studies in areas such as management and education. Monte Carlo simulations offer empirical robustness for complex or longitudinal models, while recommendations based on the type of estimator, such as ML or WLSMV, highlight the need to align sample size with data characteristics. Finally, consideration of the structural complexity of the model emerges as a central factor: second-order, multigroup, or longitudinal models require substantially larger samples to ensure parameter stability and reliability of estimates, often above 400–600 participants, reaching 700 or more in extensive models.

Finally, some suggestions for future work are provided. Several avenues can be explored to deepen and expand the findings of this study. First, comparative empirical research between sample size estimation methods in specific application contexts is recommended, with the aim of verifying whether the general recommendations hold true in different populations and data types. Another important line of research would be to evaluate the performance of the methods in highly complex or longitudinal models, including second-order factors and multigroup structures, especially considering categorical or non-normal data, to validate and refine the proposed estimates. In addition, future studies could explore the impact of sample size on measures of validity and reliability in SEM. This type of study could assess how different estimates affect parameters, factor loadings, and fit indices in real-world scenarios. It would also be relevant to develop interactive tools or guides to assist researchers in defining ideal samples, in which various variables such as estimator type, model complexity, and magnitude of expected effects can be incorporated.

Ethics Approval

Not applicable. This study is based on a narrative review and conceptual scenario-based methodological analysis and did not involve human participants, animals, or the collection of identifiable personal data; therefore, ethics committee approval was not required.

Informed Consent

Not applicable. The study did not involve human participants or the collection of personal data.

Data Availability Statement

Data are not applicable because no new empirical dataset was created or analyzed in this study. The article is based on published literature and conceptual methodological scenarios.

AI Transparency Statement

The author did not use AI-assisted tools in the preparation of this manuscript.

Acknowledgements

Not applicable.

Conflicts of Interest

The author declares no conflicts of interest.

References

Nusair, K., & Hua, N. (2010). Comparative assessment of structural equation modeling and multiple regression

research methodologies: E-commerce context. *Tourism Management, 31*(3), 314–324. https://doi.org/10.1016/j.tourman.2009.03.010

Xiong, B., Skitmore, M.R., & Xia, B. (2015). A critical review of structural equation modeling applications in construction research. *Automation in Construction,* 49, 59–70. https://doi.org/10.1016/j.autcon.2014.09.006

Allammari, Y., Taqi, A., & El Morabit, N. (2025). Entrepreneurial orientation and SME performance: A sequential mediation analysis of market and learning orientations. *Journal of Applied Structural Equation Modeling, 9*(2), 1–30. https://doi.org/10.47263/JASEM.9(2)02

Dulaimi, M.F., Nepal, M.P., & Park, M. (2005). A hierarchical structural model of assessing innovation and project performance. *Construction Management and Economics, 23*(6), 565–577. https://doi.org/10.1080/01446190500126684

Ghardashi, F., Yaghoubi, M., Teymourzadeh, E., Bahadori, M., & Salesi, M. (2022). The innovation capability model in higher education: A structural equation modelling approach. *African Journal of Science, Technology, Innovation and Development, 15*(4), 473–481. https://doi.org/10.1080/20421338.2022.2125594

Kline, R. (2010). *Principles and Practice of Structural Equation Modeling.* Guilford Press.

Cheung, G.W., Cooper-Thomas, H.D., Lau, R.S., & Wang, L.C. (2024). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. *Asia Pacific Journal of Management, 41*(2), 745–783. https://doi.org/10.1007/s10490–023–09871-y

Newsom, J.T. (2023). *Longitudinal Structural Equation Modeling: A Comprehensive Introduction.* Routledge.

Panzeri, A., Castelnuovo, G., & Spoto, A. (2024). Assessing Discriminant Validity through Structural Equation Modeling: The Case of Eating Compulsivity. Nutrients, 16(4), 550. https://doi.org/10.3390/nu16040550

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. https://doi.org/10.1207/s15328007sem1203_2

Hox, J.J., Maas, C.J.M., & Brinkhuis, M.J.S. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica,* 64, 157–170. https://doi.org/10.1111/j.1467–9574.2009.00445.x

Nicolaou, A., & Masoner, M. (2013). Sample size requirements in structural equation models under standard conditions. *International Journal of Accounting Information Systems, 14*(4), 256–274. https://doi.org/10.1016/j.accinf.2013.11.001

Ramos-Vera, C.A. (2022). Beyond sample size estimation in clinical univariate analysis. An online calculation for structural equation modeling and network analysis on latent and observable variables. *Nutrición Hospitalaria, 38*(6), 1310. https://doi.org/10.20960/nh.03751

Wang, Y.A., & Rhemtulla, M. (2021). Power Analysis for Parameter Estimation in Structural Equation Modeling: A Discussion and Tutorial. *Advances in Methods and Practices in Psychological Science, 4*(1), 1–17. https://doi.org/10.1177/2515245920918253

Wolf, E., Harrington, K., Clark, S., & Miller, M. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educ Psychol Meas, 76*(6), 913–934. https://doi.org/10.1177/0013164413495237

Ali Memon, M., Ting, H., Cheah, J.H., Ramayah, T., Chuah, F., & Cham, T.H. (2020). Sample Size for Survey Research: Review and Recommendations. *Journal of Applied Structural Equation Modeling, 4*(2), 1–20. https://doi.org/10.47263/JASEM.4(2)01

Iacobucci, D. (2010). Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology, 20*(1), 90–98. https://doi.org/10.1016/j.jcps.2009.09.003

Kush, J.M., Konold, T.R., & Bradshaw, C.P. (2021). The Sampling Ratio in Multilevel Structural Equation Models: Considerations to Inform Study Design. *Educational and Psychological Measurement, 82*(3), 409–443. https://doi.org/10.1177/00131644211020112

Buchberger, E., Ngo, C.T., Peikert, A., Brandmaier, A., & Werkle-Bergner, M. (2024). Estimating statistical power for structural equation models in developmental cognitive science: A tutorial in R. *Behavior Research Methods,* 56, 1–18. https://doi.org/10.3758/s13428–024–02396–2

Flora, D.B., Crone, G., & Bell, S.M. (2025). Effect Size Interpretation in Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal, 32*(6), 1069–1076. https://doi.org/10.1080/10705511.2025.2459768

Graf, S., & Korn, R. (2020). A guide to Monte Carlo simulation concepts for assessment of risk-return profiles for regulatory purposes. *European Actuarial Journal,* 10, 273–293. https://doi.org/10.1007/s13385–020–00232–3

Harrison, R.L. (2010). Introduction To Monte Carlo Simulation. *AIP Conf Proc,* 1204, 17–21. https://doi.org/10.1063/1.3295638

Schamberger, T. (2023). Conducting Monte Carlo simulations with PLS-PM and other variance-based estimators for structural equation models: a tutorial using the R package cSEM. *Industrial Management & Data Systems, 123*(6), 1789–1813. https://doi.org/10.1108/IMDS-07–2022–0418

Serang, S. (2020). A Monte Carlo Test for Longitudinal Structural Equation Models in Small Samples. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(2), 165–181. https://doi.org/10.1080/10705511.2020.1756293

Maydeu-Olivares, A. (2017). Maximum Likelihood Estimation of Structural Equation Models for Continuous Data: Standard Errors and Goodness of Fit. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(3), 383–394. https://doi.org/10.1080/10705511.2016.1269606

DiStefano, C., & Morgan, G.B. (2014). A Comparison of Diagonal Weighted Least Squares Robust Estimation Techniques for Ordinal Data. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(3), 425–438. https://doi.org/10.1080/10705511.2014.915373

van Kesteren, E.J., & Oberski, D.L. (2019). Exploratory Mediation Analysis with Many Potential Mediators. *Structural Equation Modeling: A Multidisciplinary Journal, 26*(5), 710–723. https://doi.org/10.1080/10705511.2019.1588124

Koufteros, X., Babbar, S., & Kaighobadi, M. (2009). A paradigm for examining second-order factor models employing structural equation modeling. *International Journal of Production Economics, 120*(2), 633–652. https://doi.org/10.1016/j.ijpe.2009.04.010

Yang, Y., Luo, Y., & Zhang, Q. (2020). A Cautionary Note on Identification and Scaling Issues in Second-order Latent Growth Models. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(2), 302–313. https://doi.org/10.1080/10705511.2020.1747938

Kelava, A., Moosbrugger, H., Dimitruk, P., & Schermelleh-Engel, K. (2008). Multicollinearity and missing constraints: A comparison of three approaches for the analysis of latent nonlinear effects. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 4*(2), 51–66. https://doi.org/10.1027/1614–2241.4.2.51

Tarka, P. (2017). An overview of structural equation modeling: its beginnings, historical development, usefulness and controversies in the social sciences. *Quality and Quantity, 52*(1), 313–354. https://doi.org/10.1007/s11135–017–0469–8

Hammond, C. (2005). The Wider Benefits of Adult Learning: An Illustration of the Advantages of Multi-method Research. *International Journal of Social Research Methodology, 8*(3), 239–255. https://doi.org/10.1080/13645570500155037

Schuberth, F., Hubona, G., Roemer, E., Zaza, S., Schamberger, T., Chuah, F., Cepeda-Carrión, G., & Henseler, J. (2023). The choice of structural equation modeling technique matters: A commentary on Dash and Paul (2021). *Technological Forecasting and Social Change,* 194, 122665. https://doi.org/10.1016/j.techfore.2023.122665

Akobeng, A.K. (2016). Understanding type I and type II errors, statistical power and sample size. *Acta Paediatrica,* 105, 605–609. https://doi.org/10.1111/apa.13384

Darling, H.S. (2024). Statistical errors in scientific research: A narrative review. *Cancer Research, Statistics and Treatment, 7*(2), 241–249. https://doi.org/10.4103/crst.crst_283_23

Rothman, K.J. (2010). Curbing type I and type II errors. European Journal of Epidemiology, 25, 223–224. https://doi.org/10.1007/s10654–010–9437–5

McQuitty, S. (2004). Statistical power and structural equation models in business research. *Journal of Business Research, 57*(2), 175–183. https://doi.org/10.1016/S0148–2963(01)00301–0

Cole, D.A., Abitante, G., Kan, H., Liu, Q., Preacher, K.J., & Maxwell, S.E. (2025). Practical Problems Estimating and Reporting Power When Hypotheses Are Embedded in Complex Statistical Models. *Advances in Methods and Practices in Psychological Science, 8*(1), 1–17. https://doi.org/10.1177/25152459241302300

Irmer, J.P. (2024). Model-implied simulation-based power estimation for correctly specified and distributionally misspecified models: Applications to nonlinear and linear structural equation models. *Behavior Research Methods,* 56, 8955–8991. https://doi.org/10.3758/s13428–024–02507-z

Hsu, H.Y., Lin, J., Skidmore, S.T., & Kim, M. (2019). Evaluating fit indices in a multilevel latent growth curve model: A Monte Carlo study. *Behavior Research Methods,* 51, 172–194. https://doi.org/10.3758/s13428–018–1169–6

Schultzberg, M., & Muthén, B. (2017). Number of Subjects and Time Points Needed for Multilevel Time-Series Analysis: A Simulation Study of Dynamic Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(4), 495–515. https://doi.org/10.1080/10705511.2017.1392862

Montoya, A.K., & Edwards, M.C. (2020). The Poor Fit of Model Fit for Selecting Number of Factors in Exploratory Factor Analysis for Scale Evaluation. *Educational and Psychological Measurement, 81*(3), 413–440. https://doi.org/10.1177/0013164420942899

Savalei, V. (2020). Improving Fit Indices in Structural Equation Modeling with Categorical Data. *Multivariate Behavioral Research, 56*(3), 390–407. https://doi.org/10.1080/00273171.2020.1717922

Chakravorti, T., Koneru, S., & Rajtmajer, S. (2025). Reproducibility and replicability in research: What 452 professors think in Universities across the USA and India. *PLoS One, 20*(3), e0319334. https://doi.org/10.1371/journal.pone.0319334

Epskamp, S. (2019). Reproducibility and Replicability in a Fast-Paced Methodological World. *Advances in Methods and Practices in Psychological Science, 2*(2), 145–155. https://doi.org/10.1177/2515245919847421

Rahal, R.M., Hamann, H., Brohmer, H., & Pethig, F. (2022). Sharing the Recipe: Reproducibility and Replicability in Research Across Disciplines. *Research Ideas and Outcomes,* 8, e89980. https://doi.org/10.3897/rio.8.e89980

Guenther, P., Guenther, M., Ringle, C., Zaefarian, G., & Cartwright, S. (2025). PLS-SEM and reflective constructs: A response to recent criticism and a constructive path forward. *Industrial Marketing Management,* 128, 1–9. https://doi.org/10.1016/j.indmarman.2025.05.003

Yuan, K.H., Tong, X., & Zhang, Z. (2015). Bias and Efficiency for SEM With Missing Data and Auxiliary Variables: Two-Stage Robust Method Versus Two-Stage ML. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(2), 178–192. https://doi.org/10.1080/10705511.2014.935750

Методи оцінювання розміру вибірки в моделюванні структурними рівняннями

Фернанду Алмейда[1ABCD]

Університет Порту та Інститут системної та комп'ютерної інженерії, технологій і науки

Авторський вклад: A – дизайн дослідження; B – збір даних; C – статаналіз; D – підготовка рукопису; E – збір коштів

Реферат. Стаття: 11 с., 2 табл., 2 рис., 48 джерел.

Передумови. Оцінювання розміру вибірки в моделюванні структурними рівняннями (SEM) залишається дискусійним методологічним питанням. Хоча правила щодо мінімального розміру вибірки широко використовуються, вони часто недостатньо інтегрують теоретичні, статистичні та практичні міркування, особливо у складних дослідницьких дизайнах.

Мета. Це дослідження має на меті здійснити синтез і порівняльний аналіз різних підходів до оцінювання розміру вибірки в SEM, інтегруючи теоретичні засади, статистичні критерії, методологічні обмеження та контексти прикладних досліджень.

Матеріали та методи. Було застосовано мультиметодний підхід, що складався з трьох компонентів: наративного огляду методів оцінювання розміру вибірки, розгляду змодельованих сценаріїв, які відображають реалістичні умови дослідження, та порівняльного оцінювання ефективності методів у цих сценаріях.

Результати. Результати показують, що емпіричні правила є корисними для попереднього оцінювання, але можуть некоректно визначати необхідний розмір вибірки у складних моделях. Аналіз статистичної потужності забезпечує точніші та більш обґрунтовані оцінки, особливо для підтверджувальних досліджень. Симуляції Монте-Карло забезпечують надійні орієнтири для складних, лонгітюдних і багатогрупових моделей. Рекомендації, що ґрунтуються на методах оцінювання параметрів, підкреслюють важливість узгодження розміру вибірки з характеристиками даних і методами оцінювання. Складність моделі виявилася ключовим чинником, причому складні моделі SEM часто потребують 400–700 і більше учасників.

Висновки. Систематичний і контекстно орієнтований підхід до оцінювання розміру вибірки є необхідним для забезпечення надійності та методологічної строгості досліджень із використанням SEM у таких галузях, як менеджмент та освіта.

Ключові слова: методологія дослідження, моделювання структурними рівняннями, оцінювання розміру вибірки, статистична потужність, складність моделі.

Information about the Authors:

Almeida Fernando: almd@fe.up.pt; https://orcid.org/0000-0002-6758-4843; University of Porto, Institute for Systems and Computer Engineering, Technology and Science (INESC TEC), Porto, Rua Dr. Roberto Frias s/n. 4200-465 Porto, Portugal.

Cite this article as: Almeida, F. (2026). Methods for Estimating Sample Size in Structural Equation Modeling. *Journal of Learning Theory and Methodology, 7*(2), 80-90. https://doi.org/10.17309/jltm.2026.7.2.03

Received: 06.06.2026. Accepted: 27.06.2026. Published: 13.08.2026