Review

Structural Equation Modeling as a Marketing Research Tool: A Guideline for SEM Users About Critical Issues and Problematic Practices

Umut ÜNAL[1]

[1]Istanbul Medeniyet University, Department of Management; umutunalesp52@gmail.com
Orcid: 0000-0001-7723-6343[1]
*Correspondence: umutunalesp52@gmail.com

Abstract: Structural equation modeling (SEM) is a very powerful multivariate statistical technique that has increasingly been used in social sciences, particularly in marketing. As a consequence of the widespread use of this contemporary analysis method, several issues that SEM users face have become a matter of concern, which are discussed thoroughly in SEM literature. This paper aims to conduct an extensive review of these issues by benefitting from the previous review works, broaden the research criteria by bringing together the issues that are separately addressed in those previous studies, and make an empirical analysis to demonstrate how well these problems are dealt with. Along with the problematic practices identified, the solutions suggested in the literature are presented. By that, this study serves as a basic guideline for SEM users.

Keywords: Structural equation modeling, Critical review, Guideline, Mispractices

1. Introduction

In social and behavioral sciences, the theories become more and more complex, necessitating composite relationships between variables to be assessed in a way that the first-generation statistical tools cannot operate [1], as a result of which the use of multivariate statistical techniques has grown considerably over the last three decades [2]–[4].

Structural Equation Modeling, or *SEM*, which can be considered a combination of factor analysis and regression, is prevalent among such second-generation statistical methods [5]. It is a multivariate statistical technique with an unrivalled ability to simultaneously test complex webs of connections between variables [6]. SEM, unlike other statistic methods, takes potential measurement errors into consideration and makes the evaluation of the model accordingly, which makes it possible to eliminate the indicators having large error terms and/or low loadings and consequently improve the quality of the constructs that form the model [7].

The paths between theoretical constructs, which are represented by latent factors, form the basis of structural equation modeling while the latent constructs are measured via observed variables (i.e. indicators). This multi-layered nature of SEM gives it a unique power to incorporate second, even third-order factors, which provides a better understanding of relationships that may not be evident prima facie [8].

The growing use of SEM in social sciences is also evident in marketing. The increase in the complexity of newly proposed theories in this discipline is reflected onto the complexity of the causal structures that are based upon these theories [9] which led the marketing researchers to apply SEM much more frequently in their studies [10]. Structural equation modeling, which was started to be used by

marketing researchers in the mid-1970s, increasingly became a method of choice for theory development and expansion beginning from the early 80s [7], [11], [12]. The developments in SEM software also contributed to the rapid expansion of this technique among the researchers. When SEM was first introduced in the 70s, the statistics software required the structural model to be specified in terms of matrices to be able to create the path diagram. However, thanks to the recent developments in software, it has become possible to specify the path diagram directly without having to create the matrices first, which is a very tedious work, especially when working on models with many constructs [5]. Nevertheless, such user-friendly software does not magically make SEM a problem-free method. On the contrary, SEM is a complex statistical technique that provides excellent solutions to very complex problems but at the same time poses complex challenges for users [6].

Many reviews in the marketing literature address the problems that researchers have regarding the use of SEM. One of the, or maybe the, most renowned of such works is the paper by Baumgartner and Homburg [11] which reviews SEM applications in four major marketing journals (the Journal of Marketing, Journal of Marketing Research, International Journal of Research in Marketing, and the Journal of Consumer Research) covering an extended period of almost two decades (1977 and 1994). Martínez-López *et al.* [10] examines the SEM-based articles published between 1995 and 2007 in the same four journals to make an update of the paper by Baumgartner and Homburg [11] and at the same time draw a period-wise comparison to demonstrate how the issues identified in the early periods of SEM have been dealt with and what new problems have arisen. Even the sheer number of articles included in these two works, 149 in the former and 472 in the latter, is evidence of the striking increase in the popularity of SEM techniques in marketing literature.

Just as there are such works examining the use of SEM in marketing area covering a particular period or articles published in specific journals (e.g., [10], [11], [13]) there are bibliometric analyses reviewing the use of SEM in marketing studies on country basis as well (e.g., [7], [14], [15]). Dogan [14] examined conference proceedings of national marketing congresses in Turkey between 1999-2017 and reviewed 91 papers that used SEM as their analysis method as an attempt to fill the gap in Turkish marketing literature regarding the issues about the use of SEM. He addresses five common problems in his study (model fit issues, omission of CFA before model testing, misuse of covariance modification indices, insufficient reporting about the constructs' reliability and validity, and mischoice of PLS-SEM over CB-SEM). Even though these are critical problems to be investigated thoroughly, some other issues discussed extensively in the marketing literature are left untouched. Therefore, in this study, it is aimed to update and continue his work by analyzing a wider set of topics central to the use of SEM applications, which are determined based on the examination of previous SEM reviews as sample size adequacy, (non)application of CFA, measurement model specification, evaluation of model fit, model respecification, reliability, validity, common method bias, normality, and mediation.

In this meta-analysis, to demonstrate the most up-to-date picture regarding the issues pinpointed in the previous studies, it was decided to focus on the most recent works. Therefore, marketing-oriented Turkish articles published in 2020 were reviewed. 120 articles using the SEM technique were detected and analyzed based on the criteria mentioned above.

In the following section, these topics and issues related to them are explained one by one.

2. Methodological Issues in The Application of SEM

SEM is a complex statistical technique, the application of which requires special attention to many critical points. However, sometimes the users of SEM fail to take into consideration certain issues that they are supposed to pay attention. To highlight these problematic applications and provide solution suggestions, many researchers reviewed the use of SEM in their field [7], [10], [11], [13]–[15]. In this study, some of the problems that were identified to be frequently faced by SEM users in these previous studies are explained.

2.1. Sample Size Adequacy

One of the most significant limitations of SEM is about its sample size requirement. It is generally acknowledged that SEM is a large-sample technique [16]. Even relatively simple models require fairly large samples, not to mention the complex structures [17]. An inadequate sample size directly affects the reliability of the results, power of analysis and generalizability of research findings [10], [13]. Therefore, sample size is one of the most crucial criteria that SEM modellers should pay attention when applying this technique.

However, even though there is a general agreement on the large sample requirement of SEM, there is no simple rule of thumb answering the question of what is a "large enough" sample size in SEM.

There are two main schools of thought that can be referred to. The first one considers the sample size in isolated terms and makes recommendations of minimum sample size independent of the number of indicators included in the model. For instance, Ding *et al.* [18] suggest a minimum of 100-150 cases. Likewise, Anderson and Gerbing [19] claim that a sample size of 150 is usually enough to obtain sufficiently convergent and proper results. Loehlin [20] recommends an absolute minimum sample size of 200 observations. The second school of thought, on the other hand, recommends determining the sample size through the number of parameters to be estimated. By many researchers this stream of though is considered more reasonable because the varying degrees of complexities of models and extra factors such as missing data make it very difficult to identify a "one-size-fits-all" sample size that will work across all SEM models [16]. In their review, Baumgartner and Homburg [11] also recommend the second school indicating that there cannot be one size that is suitable to all SEM applications. In this sense, the recommendation of *N:q* ratio by Bentler and Chou [21], which considers the ratio of number of observations to the number of free parameters as the criteria of sample size assessment and take 1:5 as acceptable and 1:10 as recommended ratio, is frequently referred to by researchers.

2.2. Measurement Model Specification

In SEM, most of the time the prime interest is on the connections between theoretical constructs which are measured by the observed variables. Therefore, how well the observed variables represent their respective unobserved (latent) variables is very crucial to the main objective of SEM analysis [5]. The substructure of the model, that is, the number of observed variables (i.e. indicators) per construct, profoundly influences the extent to which the structural model is well identified [10]. Out of the two types of constructs, i.e. single-item and multi-item constructs, Baumgartner and Homburg [11] were highly critical of the former and discouraged their use. As for the number of indicators in multi-item constructs, at least 3 indicators per construct is recommended to obtain reliable results [22].

2.3. (Non)application of Confirmatory Factor Analysis

In SEM, prior to the application of structural model by which the research hypotheses are tested, the measurement model needs to be created, and the structure of constructs and their measurement items should be identified [1]. To do this, confirmatory factor analysis (CFA) is applied by which it is tested how well the measured variables represent their respective theoretical latent constructs and how well the prespecified measurement theory matches reality as captured by data [17].

Even though CFA is a prerequisite to the hypothesis testing in SEM, Dogan [14] indicated that almost one quarter of the papers reviewed in his study lacks a CFA and directly skip to the application of structural model. Therefore, even though such an issue was not identified in other SEM reviews (e.g., [10], [11], [13]), in this paper it is included among the subjects to be examined.

2.4. Assessment of Model Fit

Goodness-of-fit indices are the indicators to assess how well the prespecified theorical model explains the data collected. Based on the model fit values, it is decided whether to accept or reject the measurement model. In case that the overall model fit is not acceptable, all the loadings, parameters and estimations within the model become null and void [17]. Therefore, the correct assessment of model fit is very crucial in SEM analysis.

Even though chi-square test is a well-known statistical method to test goodness of fit, it has a downside that it is very sensitive to sample size [1]. If the sample size is very large, which is often the case in SEM especially when working with complex models, the statistical test will almost certainly be significant and even if it actually describes the data very well, the model will always be rejected. Likewise, when the sample is small, the model will always be accepted, even if in reality it fits poorly [5]. As a solution to this problem, researchers have proposed a lot of different alternative fit indices over time, which provided SEM an "arsenal" of goodness-of-fit indices [10, pp. 129]. Among them, the most used ones are GFI, AGFI, CFI, TLI, RMSEA and SRMR. The acceptable thresholds of these indices are as follows: GFI ≥ .90, AGFI ≥ .90, CFI ≥ .90, TLI ≥ .90, SRMR ≤ .08, RMSEA ≤ .06 [1], [17], [23].

2.5. Model Respecification

When an adequate level of fit is not achieved, it is common practice to modify the model, by deleting nonsignificant parameters that are upsetting the model fit, and/or by adding new parameters, which would improve the fit if specified [5]. To make this process easier, SEM provides the user a diagnostic output, modification indices, which calculates the prospective impact of every possible covariance that is not estimated in the model on the overall fit. Adding the new covariance path suggested in the modification indices to the model reduces the overall $\chi2$ value, thus improves the model fit [17]. However, the misuse of these indices is a serious problem in SEM applications [10]. When performing model respecification by utilizing modification indices, researchers should consider the theorical basis of the changes they make [17], [19]; otherwise the modifications would amount to baseless data-driven adjustments that lack validity [24].

2.6. Reliability

Reliability refers to the assessment on degree of consistency among a set of items that measure a construct. For a scale to be reliable, its items are expected to be reflective of the same underlying latent construct and to be highly intercorrelated. In simpler terms, they are expected to be closely related as a group and vary together [17]. Establishment of construct reliability is very crucial in SEM, because only then the results achieved can be treated with confidence [6]. In marketing literature, researchers customarily apply the Cronbach's alpha to evaluate the internal consistency of measures [10]. However, this coefficient suffers from several weaknesses. First, the number of items in the test influences the accuracy of the reliability estimation [25]. Second, Cronbach's alpha operates under the assumption of equally weighted indicator loadings [22]. Third, it does not weight each individual item in the calculations [17]. Due to these limitations, Cronbach's alpha is not considered a true index for unidimensionality assessment [10]. In response to these weaknesses, more accurate construct reliability coefficients have been proposed, among which composite reliability (CR) is the most commonly applied measure. Unlike Cronbach's alpha, it is not influenced by the number of items in the scale [26], does not assume indicators to be equally loaded and weight the individual indicators based on their loadings [17]. The acceptable threshold for CR is 0.70, with each item having a minimum factor loading of 0.707 [26].

2.7. Validity

In SEM, two types of validity have to be established in order to obtain statistically admissible results. These are convergent validity and discriminant validity. Convergent validity is a measure of how closely the indicators of a scale converge, or 'load together,' on a single latent construct. In other words, it tests if the measures of a construct which are theoretically related are really related or not [17]. Convergent validity, or communality, is assessed through the calculation of average variance extracted (AVE) across all indicators linked with a particular construct. The rule of thumb for an acceptable AVE, which is the average of the squared loadings of all indicators under a particular construct, is 0.50 or higher [26].

Discriminant validity, on the other hand, measures if the indicators that theoretically should not be related are indeed not related. In other words, discriminant validity is established unless the items belonging to different constructs move very closely as if they were the members of the same group [17].

Typically, CB-SEM uses Fornell–Larcker criterion, which verifies discriminant validity if the AVE of a particular construct is higher than the variance that construct shares with each of other constructs [26]. An alternative method for assessing discriminant validity is heterotrait-monotrait ratio of correlations (HTMT), which was recently proposed by Henseler *et al.* [27] . HTMT is considered is a more precise contemporary measure of assessment. Even though it was originally recommended to be used in PLS-SEM, HTMT can be used in any SEM model regardless of the estimation method used [28]. An HTMT value below 0.90 suggests the evidence of discriminant validity [27].

2.8. Common Method Bias

Common method bias (CMB), or common method variance, is the spurious variance that can be attributed to the measurement instrument rather than to the constructs that the measures are assumed to represent [29]. Among possible causes of CMB, social factors such as implicit social desirability and acquiescence are the possibilities that are frequently mentioned in the literature [6], [30], [31]. Also the research instrument can cause CMB; for instance, the instructions on the top of a survey may lead the responses of the participants to a certain direction. As a result, the constructs end up sharing a certain amount of common variation which is not actually related to the network of causal relationships in the model tested [30]. CMB is a very serious phenomenon that needs to be checked because it can bias the reliability and validity of measures, which will cast suspicion on the correctness of the analysis results obtained, and can lead to misjudgements about the hypotheses tested in a research model because it may inflate or deflate the estimates between two constructs [31]. Nevertheless, despite the widespread acknowledgement about the problems that it poses, a surprising volume of studies do not check the existence of CMB in their data [30], [31]. To make sure if the data is contaminated with CMB or not, the most frequently used method is to apply Harman's single factor test, which measures if a single factor accounts for the majority of the covariance among the constructs. If the cumulative variance extracted by one factor is lower than the threshold of 50 per cent it can be concluded that CMB is not present in the data [32]. Some researchers show an inclination to test the reliability and validity of the constructs but skip measuring CMB. This is an important problem to be addressed because models may have acceptable convergent and discriminant validity values but still be contaminated by common method bias [30].

2.9. *Normality*

Normality is another concern involving SEM analysis. Most estimation methods in SEM, including the most popular ones (maximum likelihood estimation and general least squares), operates under the assumption of normal distribution of the data. Violation of this rule may lead to distorted goodness-of-fit measures and undervalued standard error terms [33]. Therefore, researchers need to check the multivariate normality of their data and, if necessary, apply remedies to account for non-normality [6]. To test the normality of data, skewness and kurtosis values need to be calculated. While skewness is used to describe the balance of the distribution; i.e. if the data are distributed symmetrically or accumulated on side (left or right), kurtosis is the measure of peakedness or flatness of distribution [17]. In the case that the data show non-normal distribution, bootstrapping is a method that can be resorted to. Bootstrapping, which refers to the regeneration of sampling distributions via resampling without replacing the original data, can be used when fitting covariance structures to data with non-normal distribution [34]. By using bootstrapped samples, researchers can reach estimations with accurate significance levels and appropriate standard errors from non-normal data [35]. Another way to account for data non-normality is to use alternative non-conventional estimations techniques such as weighted least squares (WLS), general weighted least squares (GWLS), and elliptical reweighted least squares (ERLS), which are, unlike maximum likelihood (ML), the most frequently used estimation technique, have less stringent requirements regarding data distribution. However, this method is discouraged because changing the estimation technique may violate the theoretical logic underpinning the original dataset [6].

2.10. Mediation

In SEM, researchers can make many complex and comprehensive analyses which are very difficult to do with first generation statistical methods. One of such calculations is mediation analysis [7], which refers to the testing of a hypothesized causal chain in which one variable affects a second variable which, in turn, affects a third variable. The variable in the middle of this causal chain, that is, the mediator, intervenes the relationship between the independent and the dependent variables [17]. The result of the mediation analysis shows the indirect effect of the independent varible, i.e predictor, on the dependent variable, i.e outcome. When testing this indirect effect, researchers resort to two general methods, which are bootstrapping and Monte Carlo method [36].

3. Descriptive Results and Findings

3.1. Sample Size

The articles reviewed in this study show that generally researchers are aware of the large sample size requirement of SEM analysis. In these 120 articles, only 3 of them have a sample size smaller than 200. Nevertheless, as explained in the previous section, it is widely recommended to assess the adequacy of sample size not in terms of absolute numbers, rather based on the complexity of the model, that is, the number of parameters to be estimated. Therefore, the sample sizes of these articles were assessed based on the of N:q ratio by proposed by Bentler and Chou [21], taking 1:5 as the threshold for an acceptable sample size.

In order to calculate the free parameters in a model, in addition to the structure among the latent variables, the number of observed variables need to be known as well. However, 6 articles (5%) share only the path diagrams between the latent constructs without any information regarding the number of observed variables, therefore no assessment could be done whether the sample size is large enough or not. This is a critical issue that needs special attention. Considering the widely acknowledged sample size requirements of SEM in the literature, researchers are recommended to report at least the number, if not the full detail, of items of each construct.

When the remaining 114 articles are analyzed, it is seen that 21 of them (18%) have insufficient sample size. Actually, the median sample size of those articles is 309. So, most of them seemingly have quite large sample sizes but the complexity of their models makes these samples "large but not large enough". The rest, 93 articles (77%), are assessed to have sufficient sample sizes.

Table 1. Sample Size Adequacy

	Frequency	Mean	Percentage
Articles with adequate sample size	93	408	77%
Articles with insufficient sample size	21	309	18%
Unsuitable for interpretation	6	455	5%
Total	120	403	100%

During the sample size assessments, another issue that is not identified in previous works came into sight, which is the misuse of Kaiser-Meyer-Olkin (KMO) and Bartlett's tests. KMO and Barlett's test of sphericity are used to determine whether the sample is suited for factor analysis by measuring the proportion of variance and correlations among variables, that is, if the data is suitable to form factors, these factors can explain each other and related to each other [17]. However, it is detected that in some articles these values are used to check the sample size adequacy. This is a critical misapplication because these tests should be applied to see if the quality, not quantity, of the data is adequate for the factor analysis.

Out of 120 articles review in this study, 69 did not use KMO and Bartlett's tests. 10 articles of the remaining 51 misused these tests and take the KMO and Bartlett's measures to conclude if the sample size of their study is adequate. This frequency corresponds to almost 20% which is not a negligible ratio.

Therefore, the researchers are advised to pay attention to this issue and verify the sample size of their study is large enough by using the criteria acknowledged in the literature.

3.2. (Non)application of CFA

The proper application of SEM technique requires a confirmatory factor analysis (CFA) prior to the hypothesis testing. In other words, first a measurement model is created to see how well the measured variables represent the number of constructs specified in the model [17]. After it is verified that factor structure fits the data, then the structural model is finalized. Even though such a problem is not identified in other leading SEM review studies (see [7], [10], [11], [13], [15]) Dogan [14] indicated that in some works researchers skip CFA and proceed directly to the hypothesis testing by building a structural model. Out of 120 articles analyzed in this study, in 25 of them (21%) CFA is not performed. This is a clear violation of the universally acknowledged two-step approach of SEM [19] that the researchers are strongly advised to follow.

3.3. Measurement Model Specification

The extent how well the observed variables represent their respective unobserved variables directly influences the quality of model specification [10]. Thus, in the previous SEM reviews, the ratio between observed/latent variables was examined to see if each construct has ideally sufficient indicators under them (see [10], [11], [13]). The median ratio of observed/latent variables for the articles reviewed in this study is 4.1, which is more than enough based on the recommendation of "at least three indicators per construct" by Baumgartner and Homburg [11].

Secondly, the use of single and multi-item constructs is analyzed. Measurement of constructs by only one item is strictly discouraged and instead the use of multi-item measures is conventionally recommended in the literature [11], [37]. Excluding the 6 articles that didn't explicitly give information regarding the number of measurement items used in the analysis, most of the remaining articles (112) used multi-item constructs. Only 2 articles used single item constructs, which corresponds to less than 2% of the works reviewed in this study. However, even though it was generally refrained from using single-item constructs, the use of constructs measured by only two indicators is not equally uncommon. Despite the use of only 2 items to measure a construct is generally opposed [11], [22], 29 articles (24%) under the scope of this review include two-item constructs in their models.

Table 2. Construct types

	Frequency	Percentage
Constructs with one item	2	2%
Constructs with two items	29	24%
Constructs with three or more items	83	69%
Non-specified	6	5%
Total	120	100%

3.4. Evaluation of Model Fit

In SEM, reporting the model fit results is extremely crucial because through the model fit indices it can be understood how well the prespecified theoretical model fits the data collected. If the model does not fit, in other words, does not explain the data, then further analysis is meaningless [5], [22]. To test the fit of the model, there are several goodness of fit indices available to researchers. The most widely used of these indices are chi-square test, incremental fit index (IFI), comperative fit index (CFI), normed fit index (NFI), relative fit index (RFI) and root mean square error of approximation (RMSEA) [10]. The indices used in the articles reviewed and their frequency are as follows:

Table 3. Use of Model Fit Indices

Model Fit Index	Frequency	Percentage
χ^2/df	93	87%
CFI	92	86%
RMSEA	90	84%
GFI	77	72%
NFI	64	60%
AGFI	46	43%
SRMR	43	40%
IFI	33	31%
TLI	20	19%
RFI	9	8%

As it can be seen in the Table 3, chi-square test, CFI, RMSEA, CFI, GFI and NFI are used most of the articles.

Out of 120 articles reviewed, 13 of them (11%) do not share any information regarding the model fit. This is a misapplication that should definitely be abstained from because if it is not known whether the model fit is adequate or not, the analysis results lose their credibility and validity.

During the examination of the use of model fit indices, another problem that has not been explicitly addressed in the previous works is identified. Even though 107 articles reported the fit values of their models, 44 of them (36%) does not indicate the goodness of fit values of both the measurement and the structural models. The proper application of SEM requires the researchers to measure the fit of both models [38] but in these 44 articles the model fit values of either the measurement or the structural model are calculated. SEM users should avoid this kind of a mispractice and report the fit values of both models.

Table 4. Reporting of Model Fit

	Frequency	Percentage
Both measurement and structural models	63	53%
Only structural model	28	23%
Only measurement model	16	13%
Not used	13	11%

3.5. *Model Respecification*

In SEM, once the measurement model is built and the model fit is calculated, adjustments on the model can be made in order to improve model fit. Modification indices that are suggested by the SEM software make this respecification process very easy. However, previous studies indicate that sometimes the modifications made by the researchers may lack theoretical plausibility even if technically possible [10], [14]. Out of 120 articles review in this study, 40 of them (33%) report that they made use of modification indices in order to improve the fit of the model. In these 40, most of them (31) utilized the modification indices properly. For 4 articles, it cannot be interpreted whether the use of modification indices is proper or not because in these studies, even though the authors indicate that they used modification indices, no further detail is presented.

For the remaining 5 articles, modification indices are assessed to be employed in an inappropriate way. The reasons leading to this assessment can be divided into two: misuse and overuse.

It is needless to say that SEM software produce the modification indices with a data driven approach. It is researchers' responsibility to judge if these suggestions are rational and well-reasoned on a theorical basis [17], [19], [24]. In 2 articles, the researchers draw correlations between the error terms belonging to different constructs. Even though such a modification may have a favorable impact on the overall fit of the model, there cannot be any theoretical foundation to correlate two error terms under different latent factors [39].

In 3 articles, there is an overuse of modification indices. Many authors advise the researchers to be cautious in using modification indices and recommend that such modifications should be kept at a minimum level [33], [40], [41]. Despite this call for caution, in these 3 articles correlations are drawn between almost all the error terms under the same construct. Such an application may improve the model fit substantially, but it is deemed problematic in the literature [39].

Table 5. Model Respecification

	Frequency
No model respecification reported	80
Model respecification reported	40
Appropriate respecification	31
Inappropriate respecification	5
Misuse of modification indices	2
Overuse of modification indices	3
Non-specified	4

3.6. Reliability

As explained in the previous section, to test reliability of the constructs in SEM, two main measures are used: Cronbach's alpha and composite reliability. In this review, it is examined in how many articles, and with which measure the reliability test is employed. Out of 120 studies, only two does not report any reliability result at all, which means that more than 98% of the articles conducted reliability test. While in 60 articles (50%) only Cronbach's alpha is used, only 5 articles (4%) use CR alone. The remaining 53 articles (44%) include both alpha and CR values in their analyses.

Table 6. Reliability measures

	Frequency
Reliability not measured	2
Reliability measured	118
Cronbach's alpha alone	60
Composite reliability (CR) alone	5
Cronbach's alpha and CR together	53

Cronbach's alpha is considered to be most commonly used measure of scale reliability in SEM applications [6] and this study points to the same finding. More than 94% of the articles reviewed use Cronbach's alpha either alone or together with CR. However, due to the limitations of alpha value explained in the previous section, it is not recommended to use only this measure. Even Cronbach himself expresses that it would be safer to use alpha coefficient along with other successor procedures rather than alone [42]. Therefore, the researchers are recommended to use both Cronbach's alpha and CR to determine the reliability of their constructs.

3.7. Validity

In this review, the articles are examined if they include proper controls of convergent and discriminant validity of the constructs. Out of these two validity types, convergent validity is much more commonly

measured. 74 articles (62%) in 120 test the convergence of their indicators. When it comes to discriminant validity, it is measured in much fewer studies. Only in 45 articles (38%) discriminant validity is analyzed (33 follow only Fornell and Larcker procedure; 11 employ both Fornell & Larcker procedure and the HTMT method; and only 1 use HTMT method alone). In 46 articles (38%), neither convergent nor discriminant validity is measured. This result indicates an alarming situation because just like reliability, validity is also a prerequisite in SEM application. Therefore, it is surprising to see that construct validity is ignored by such a big portion of these studies even though reliability is addressed in almost all the articles reviewed.

Table 7. Validity measures

	Frequency
Validity not measured	46
Validity measured	74
Convergent validity	74
Discriminant validity	45

3.8. Common Method Bias

Despite its critical importance, in several previous SEM review works CMB has not been scrutinized (see [10], [11], [13]–[15]). To fill this gap in the literature and set an example for the future SEM review studies, the articles are examined to see if CMB is addressed or not. Out of 120 articles, only 9 of them (8%) ascertain that their data is not contaminated with CMB. All these 9 articles use Harman's single factor test as the measurement method.
SEM researchers collect their data very frequently through surveys, which are always under a certain risk of being affected by CMB [30], [31]. Therefore, to establish credibility of their analyses, it needs to become a common practice for SEM user to ensure that CMB is not a major issue in their dataset and take preventive countermeasures against CMB such as guaranteed anonymity and confidentiality, reverse coding of construct items, pretesting for item wording refinement and improvement [43].

3.9. Normality

Even though normal data distribution is one of the assumptions of SEM, previous review studies show that users of SEM fail to report the normality results of their data more often than not [10], [13]. Therefore, in this study reporting of data distribution is examined. In 23 articles (19%) the partial least squares estimation technique (PLS-SEM) was used. Out of these 23 articles, only 5 report normality measures (4 of them report kurtosis and skewness measures while 1 conducted Kolmogorov-Smirnov test). In the literature, there are two views regarding the normality requirement of PLS. The traditional view asserts that PLS-SEM, in contrast with CB-SEM, does not need the data to be distributed normally. Therefore, PLS-SEM based studies do not need to conduct a normality test [8]. The modern view, on the other hand, claims that PLS-SEM does not differ from CB-SEM regarding normality assumption. The non-parametric technique, bootstrapping, that PLS-SEM applies can equally be applied by other SEM techniques [44]. Therefore, since there is no unity in the literature agreeing that PLS-SEM based studies do not need to report the distribution results of their data, it can be seen a safer way to perform the normality test regardless of the estimation method and refer to bootstrapping in case of non-normal distribution as a remedy. 66 articles (56%) of the remaining 97 do not examine the normality of their data. Only in 31 articles (26%) measures of normality are reported (27 used kurtosis and skewness values; 7 conducted Kolmonogov-Smirnov test; and 1 employed Shapiro–Wilk test). In all the studies that report normality results, there is no non-normal distribution issue expect for 3 articles, all of which utilize bootstrapping as a solution.

Table 8. Normality measures

	Frequency	Estimation Method

		CB-SEM	PLS-SEM
Normality not measured	84	66	18
Normality measured	36	31	5

3.10. Mediation

Researchers widely use SEM applications when conducting mediation analyses. In literature, there are many studies recommending the use of bootstrapping to define the confidence intervals for mediation effects [45], [46]. Therefore, in this study, it is examined how many articles test the mediating effect of a construct and how many of these articles make use of bootstrapping.

Out of 120 articles reviewed, 29 of them (24%) have analyses involving a mediating effect. In these 29 studies, 10 use bootstrapping technique. Also in one article Sobel test was applied. Therefore, only 11 articles (38%) out of 29 employed a statistical method designed to be used in mediation analyses.

4. Conclusion and Implications

In this study, 10 topics central to the use of SEM are examined by reviewing 120 SEM-based articles which are all published in 2020. The selection of these recently published articles ensures that all the issues detected in this review are current, not outdated, problems in SEM applications. The issues detected in previous SEM reviews form the backdrop to this study. Also some critical problems discussed in many works in the literature but were not included in those leading SEM reviews are included within the scope of this study.

SEM is a very effective statistical method frequently used for theory testing and development. As a tool that is used by researchers very commonly, it has a substantial role in the generation of new knowledge in many disciplines. However, only when it is applied properly, the information obtained through SEM analysis can be treated with confidence. Therefore, researchers should pay the utmost attention to the problematic practices, misuses and critiques regarding SEM applications when employing this technique.

Even though the descriptive statistics explain the practices regarding the use of SEM in Turkish marketing literature, the issues identified in this study are not limited to one discipline or country. Therefore, this study serves as a guideline to all SEM users about the most common problems faced in SEM applications, wrongful practices that should be avoided, exercises that needs to be executed and remedies that may be applied.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1]   R. P. Bagozzi and Y. Yi, "Specification, evaluation, and interpretation of structural equation models," *Journal of the Academy of Marketing Science*, vol. 40, no. 1, Jan. 2012, doi: 10.1007/s11747-011-0278-x.

[2]   W. W. Chin, R. A. Peterson, and S. P. Brown, "Structural Equation Modeling in Marketing: Some Practical Reminders," *Journal of Marketing Theory and Practice*, vol. 16, no. 4, Sep. 2008, doi: 10.2753/MTP1069-6679160402.

[3]   M. Sarstedt, C. M. Ringle, and J. F. Hair, "PLS-SEM: Looking Back and Moving Forward," *Long Range Planning*, vol. 47, no. 3, pp. 132–137, Jun. 2014, doi: 10.1016/J.LRP.2014.02.008.

[4]   R. Weston and P. A. Gore, "A Brief Guide to Structural Equation Modeling," *The Counseling Psychologist*, vol. 34, no. 5, Sep. 2006, doi: 10.1177/0011000006286345.

[5]   J. J. Hox and T. M. Bechger, "Introduction Structural Equation Modeling An Introduction to Structural Equation Modeling 1 What is Structural Equation Modeling?," 1998.

[6]   G. T. M. Hult *et al.*, "An Assessment of the Use of Structural Equation Modeling in International Business Research," *Research Methodology in Strategy and Management*, vol. 3. pp. 385–415, 2006. doi: 10.1016/S1479-8387(06)03012-8.

[7] J. F. Hair Jr., M. L. D. da S. Gabriel, and V. K. Patel, "AMOS Covariance-Based Structural Equation Modeling (CB-SEM): Guidelines on Its Application as a Marketing Research Tool," *Brazilian Journal of Marketing*, vol. 13, no. 2, pp. 44–55, May 2014, doi: 10.5585/remark.v13i2.2718.

[8] C. B. Astrachan, V. K. Patel, and G. Wanzenried, "A comparative study of CB-SEM and PLS-SEM for theory development in family firm research," *Journal of Family Business Strategy*, vol. 5, no. 1, pp. 116–128, Mar. 2014, doi: 10.1016/J.JFBS.2013.12.002.

[9] C. L. Shook, D. J. Ketchen, G. T. M. Hult, and K. M. Kacmar, "An assessment of the use of structural equation modeling in strategic management research," *Strategic Management Journal*, vol. 25, no. 4, Apr. 2004, doi: 10.1002/smj.385.

[10] F. J. Martínez-López, J. C. Gázquez-Abad, and C. M. P. Sousa, "Structural equation modelling in marketing and business research: Critical issues and practical recommendations," *European Journal of Marketing*, vol. 47, no. 1, pp. 115–152, Feb. 2013, doi: 10.1108/03090561311285484.

[11] H. Baumgartner and C. Homburg, "Applications of structural equation modeling in marketing and consumer research: A review," 1996.

[12] P. M. Bentler, "Multivariate Analysis with Latent Variables: Causal Modeling," *Annual Review of Psychology*, vol. 31, no. 1, Jan. 1980, doi: 10.1146/annurev.ps.31.020180.002223.

[13] Y. Koubaa, R. S. Tabbane, and R. C. Jallouli, "On the use of structural equation modeling in marketing image research," *Asia Pacific Journal of Marketing and Logistics*, vol. 26, no. 2, pp. 315–338, 2014, doi: 10.1108/APJML-10-2013-0113.

[14] V. Doğan, "PAZARLAMA ARAŞTIRMACILARININ YAPISAL EŞİTLİK MODELİ ANALİZİ UYGULAMALARI: SORUNLAR VE ÖNERİLER," *Journal of Administrative Sciences*, vol. 16, no. 32, pp. 201–230, 2018.

[15] D. Frías-Navarro and M. P. Soler, "Exploratory factor analysis (EFA) in consumer behavior and marketing research," *Suma Psicológica*, vol. 19, pp. 47–58, 2012.

[16] Kline Rex B., *Principles and Practice of Structural Equation Modeling*, 4th ed. New York: The Guilford Press, 2016.

[17] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, 8th ed. Hampshire, UK: Cengage Learning EMEA, 2019. [Online]. Available: www.cengage.com/highered

[18] L. Ding, W. F. Velicer, and L. L. Harlow, "Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices," *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 2, no. 2, Jan. 1995, doi: 10.1080/10705519509540000.

[19] J. C. Anderson and D. W. Gerbing, "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," 1988.

[20] J. C. Loehlin, *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates, 1998.

[21] P. M. BENTLER and C.-P. CHOU, "Practical Issues in Structural Modeling," *Sociological Methods & Research*, vol. 16, no. 1, Aug. 1987, doi: 10.1177/0049124187016001004.

[22] K. A. Bollen, *Structural Equations with Latent Variables*. New York, NY: Wiley Interscience, 1989.

[23] L. Hu and P. M. Bentler, "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives," *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 6, no. 1, Jan. 1999, doi: 10.1080/10705519909540118.

[24] R. MacCallum, "Specification searches in covariance structure modeling.," *Psychological Bulletin*, vol. 100, no. 1, 1986, doi: 10.1037/0033-2909.100.1.107.

[25] M. S. Garver and J. T. Mentzer, "Logistics research methods: Employing structural equation modeling to test for construct validity," *Journal of Business Logistics*, vol. 20, no. 1, pp. 33–57, 1999.

[26] C. Fornell and D. F. Larcker, "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, vol. 18, no. 1, Feb. 1981, doi: 10.1177/002224378101800104.

[27] J. Henseler, C. M. Ringle, and M. Sarstedt, "A new criterion for assessing discriminant validity in variance-based structural equation modeling," *Journal of the Academy of Marketing Science*, vol. 43, no. 1, Jan. 2015, doi: 10.1007/s11747-014-0403-8.

[28] J. F. Hair Jr., L. M. Matthews, R. L. Matthews, and M. Sarstedt, "PLS-SEM or CB-SEM: updated guidelines on which method to use," *International Journal of Multivariate Data Analysis*, vol. 1, no. 2, 2017, doi: 10.1504/IJMDA.2017.087624.

[29] P. M. Podsakoff, S. B. MacKenzie, J. Y. Lee, and N. P. Podsakoff, "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies," *Journal of Applied Psychology*, vol. 88, no. 5. pp. 879–903, Oct. 2003. doi: 10.1037/0021-9010.88.5.879.

[30] N. Kock, "Common method bias in PLS-SEM: A full collinearity assessment approach," 2015.

[31] P. J. Jordan and A. C. Troth, "Common method bias in applied settings: The dilemma of researching in organizations," *Australian Journal of Management*, vol. 45, no. 1, pp. 3–14, Feb. 2020, doi: 10.1177/0312896219871976.

[32] P. M. Podsakoff and D. W. Organ, "Self-Reports in Organizational Research: Problems and Prospects," *Journal of Management*, vol. 12, no. 4, Dec. 1986, doi: 10.1177/014920638601200408.

[33] R. C. MacCallum, M. Roznowski, and L. B. Necowitz, "Model modifications in covariance structure analysis: The problem of capitalization on chance.," *Psychological Bulletin*, vol. 111, no. 3, 1992, doi: 10.1037/0033-2909.111.3.490.

[34] K. A. Bollen and R. A. Stine, "Bootstrapping Goodness-of-Fit Measures in Structural Equation Models," *Sociological Methods & Research*, vol. 21, no. 2, Nov. 1992, doi: 10.1177/0049124192021002004.

[35] A. J. Tomarken and N. G. Waller, "Structural equation modeling: Strengths, limitations, and misconceptions," *Annual Review of Clinical Psychology*, vol. 1. pp. 31–65, 2005. doi: 10.1146/annurev.clinpsy.1.102803.144239.

[36] D. Tofighi and D. P. MacKinnon, "Monte Carlo Confidence Intervals for Complex Functions of Indirect Effects," *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 23, no. 2, Mar. 2016, doi: 10.1080/10705511.2015.1057284.

[37] J.-B. E. M. Steenkamp and H. Baumgartner, "On the use of structural equation models for marketing modeling," *International Journal of Research in Marketing*, vol. 17, no. 2–3, Sep. 2000, doi: 10.1016/S0167-8116(00)00016-1.

[38] J. B. Schreiber, A. Nora, F. K. Stage, E. A. Barlow, and J. King, "Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review," *The Journal of Educational Research*, vol. 99, no. 6, Jul. 2006, doi: 10.3200/JOER.99.6.323-338.

[39] R. Hermida, "The problem of allowing correlated errors in structural equation modeling: concerns and considerations," *Computational Methods in Social Sciences*, vol. 3, no. 1, pp. 5–17, 2015.

[40] A. J. Tomarken and N. G. Waller, "Potential problems with 'well fitting' models.," *Journal of Abnormal Psychology*, vol. 112, no. 4, 2003, doi: 10.1037/0021-843X.112.4.578.

[41] R. Shah and S. M. Goldstein, "Use of structural equation modeling in operations management research: Looking back and forward," *Journal of Operations Management*, vol. 24, no. 2, pp. 148–169, Jan. 2006, doi: 10.1016/J.JOM.2005.05.001.

[42] L. J. Cronbach and R. J. Shavelson, "My Current Thoughts on Coefficient Alpha and Successor Procedures," *Educational and Psychological Measurement*, vol. 64, no. 3, Jun. 2004, doi: 10.1177/0013164404266386.

[43] I. Rodríguez-Ardura and A. Meseguer-Artola, "Editorial: How to Prevent, Detect and Control Common Method Variance in Electronic Commerce Research," *Journal of theoretical and applied electronic commerce research*, vol. 15, no. 2, 2020, doi: 10.4067/S0718-18762020000200101.

[44] J. Henseler, G. Hubona, and P. A. Ray, "Using PLS path modeling in new technology research: updated guidelines," *Industrial Management & Data Systems*, vol. 116, no. 1, Feb. 2016, doi: 10.1108/IMDS-09-2015-0382.

[45] G. W. Cheung and R. S. Lau, "Testing Mediation and Suppression Effects of Latent Variables," *Organizational Research Methods*, vol. 11, no. 2, Apr. 2008, doi: 10.1177/1094428107300343.

[46] D. P. MacKinnon, C. M. Lockwood, and J. Williams, "Confidence Limits for the Indirect Effect: Distribution of the Product and Resampling Methods," *Multivariate Behavioral Research*, vol. 39, no. 1, Jan. 2004, doi: 10.1207/s15327906mbr3901_4.