The advancement of the built environment research through employment of structural equation modeling (SEM)

S Wasilah[1] **and T Fahmyddin** [2]
[1] Architecture Department, Islamic State of The University of Alauddin, Makassar. Jl. SultanAlauddin No. 36 Samata Kab. Gowa (Kampus 2). Indonesia.

[2] Architecture Department, Islamic State of The University of Alauddin, Makassar. Jl. SultanAlauddin No. 36 Samata Kab. Gowa (Kampus 2). Indonesia.

E-mail: wasilah@uin-alauddin.ac.id

Abstract. The employment of structural equation modeling (SEM) in research has taken an increasing attention in among researchers in built environment. There is a gap to understand the attributes, application, and importance of this approach in data analysis in built environment study. This paper intends to provide fundamental comprehension of SEM method in data analysis, unveiling attributes, employment and significance and bestow cases to assess associations amongst variables and constructs. The study uses some main literature to grasp the essence of SEM regarding with built environment research. The better acknowledgment of this analytical tool may assist the researcher in the built environment to analyze data under complex research questions and to test multivariate models in a single study.

1. Introduction

The employment of structural equation modeling (SEM) in research has taken considerable numbers in the various field of disciplines and a greater application among researchers in built environment disciplines [1-4]. SEM comprises a statistical technique to test hypotheses on the relationships among observed and latent variables [2]. SEM represents**,** estimates and test a hypothetical network of a linear relationship among observable or unobservable variables. SEM combines factor analysis and multiple regression analysis to analyze the structural relationship between measured variables and latent constructs [5]. Substantially, SEM examines the relationship between constructs or variables of interest in the research.

Various studies employed SEM approach for modeling and analyzing research problems in various disciplines such as psychological, social science, education, etc. [5-8]. The application of SEM has provided the benefits to estimating, assessing and presenting the model in a causal path model to confirm the hypothesi**z**ed relationships among the constructs of interest [9,10]. SEM provides the opportunity to the modify in improving the tested model while the model is not fit to the data in the process of the empirical model tested against the hypothesized model for the goodness of fit. Although SEM take a high consideration in various disciplines, but the discussion of application in amongst built environment disciplines is rare. This article aims at offering rudimentary knowledge of structural equation modeling (SEM) approach in data analysis, attributes, and application opportunity in built environment.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

2. SEM Overview

Many studies found that the use of SEM to be more appropriate to address a variety of research questions [11-13]. SEM comprises measurement model and a structural model [2]. The measurement model describes the relationships between observed variables and latent (unobserved) variables. It also measures the hypothesized latent (unobserved) variables. Using confirmatory factor analysis (CFA), the measurement model can examine how well the observed variables combine to identify underlying hypothesized constructs.

The latent variables must have at least three variables or indicators for measurement, see Fig.1. It aims to secure the reliability of the observed indicators and to guarantee the models can have few errors [14,15]. In further, a study can decide on the observed indicators to define the latent factors in the measurement model. The degree to which a latent variable is accurately determined depending on how strongly linked the observed indicators are. Model misspecification in the hypothesized relationships among variables appears while an indicator is weakly related to others and this creates in a deficient acknowledgment of the latent variable [4].

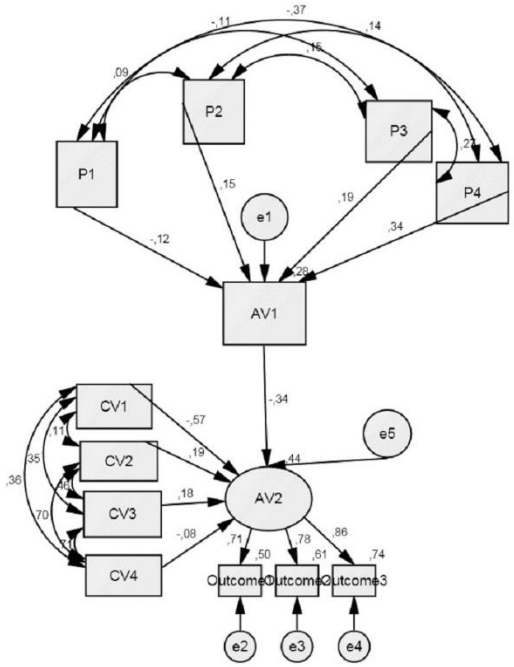


Figure 1. Example of Measurement Model

According to the figure 1, there are two latent factors of AV1 and AV2 that are estimated by a different number of observed variables. Rectangle shapes represent the observed variables while the oval shapes show the latent variables. The straight lines with an arrow at the end represents a hypothesized effect one variable has on another. The ovals shape indicators of each observed variables represent the measurement errors (residuals) indicated with e1 to e4. On the other hand, the structural model deals with the nature and magnitude of the interrelationships among constructs [16]. This model shows the relations between the latent variables which are the hypothesized to be measured.

3. Assessment of model

3.1. Assessment of measurement model

The statistical method for analysis measurement model is Confirmatory Factor Analysis (CFA). CFA validates measurement models for latent constructs with the requirement of study has theories of the underlying latent variable structure [17]. Latent constructs are predicted by the dimensions that

influenced by several indicators. Evaluating the measurement model aims to determine the validity and reliability of the measured used to represent the construct of interest [17]. Building the measurement model consist of three steps. These are checking the significance of the parameter estimates, checking the model fit, and assessing the validation [17-19]. The calculation of critical ratio and p values from statistical significance of the indicators regression weight is the first dimension in measurement model development. If the critical ratio for a regression weight is higher than +1.96 or lower than -1.96, or have a p value less than .05, that means that particular indicator has a statistically significant influence on the latent construct [17,20].

The second dimension in development is assessing the fitness by using some indexes. SEM calculates the measurement model using the fit indexes (see Table 1), as follow:

1. Model chi-square ($\chi 2$). The Chi-Square value measures the fitness of model and, to quantify the degree of discrepancy between the sample and fitted covariances matrices [20,21]. A good model fit must meet result at a 0.05 threshold [22];
2. Root mean square error of approximation (RMSEA). The RMSEA confirms how well the model, with unknown but optimally chosen parameter estimates, fit the population's covariance matrix [17]. The index of RMSEA that range between 0.08 to 0.10 confirms mediocre fit while below 0.08 shows a good fit model [23];
3. Goodness of fit statistic (GFI) and the adjusted goodness of fit statistic (AGFI). These measures are alternative to the Chi-Square test. They calculate the proportion of variance through the assessment of population covariance [24]. For the GFI, an accepted an omnibus cut-off point is 0.90 [25];
4. Root mean square residual (RMR) and standardi**z**ed root mean square residual (SRMR). The RMR and the SRMR are defined as the square root of the difference between the residuals of the example covariance matrix and the hypothesized covariance model. Values for the SRMR is from 0 to 1.0 ruling that fit model must have values less than 0.05 [17];
5. Normed-fit index (NFI). Normed-fit index (NFI) calculates the model through comparing the $\chi 2$ value of the model to the $\chi 2$ of the null model. Values NFI range between 0 and 1 with recommending values greater than 0.90 representing "a good fit" model [26];
6. CFI (Comparative fit index). This index is revision of the NFI for small sample size case [23, 27, 28]. CFI assumes that all latent variables do not correlate and compares the sample covariance matrix with a null model. CFI values range between 0.0 and 1.0 with values closer to 1.0 indicating good fit. The value of CFI $\geq$ 0.95 is acknowledged as "a good fit" model [19].

Table 1. The Index of goodness of fit

Fit Index	Cut-off criteria	Sources
Statistic Chi Square (χ^2)	Fail to reject H0	Barret (2007), L Hu & Bentler (1999)
Goodness of Fit Index (GFI)	$\geq$ .0.90	Tabachnick & Fidell (1989), Shevlin & Miles (1998).
Standardized Root Mean Square Residual (SRMR)	$\leq$ .05	Byrne (1998)
Root Mean Square Error of Approximation (RMSEA)	0.08 to 0.10 = mediocre fit and $\geq$ 0.08 = good fit	MacCallum et al. (1996).
Non-Normed Fit Index (NNFI)	$\geq$ .90	
Normed Fit Index (NFI)	$\geq$ .90	Bentler and Bonnet (1980)

Incremental Fit Index (IFI)	≥ .90	
Comparative Fit Index (CFI)	≥ 0.90	Hu and Bentler (1999)

The assessment of reliability is the third dimension in the measurement model development. It aims to test to what extent a measuring instrument can give the same relative output for the same cases or models. According to Variance Extract/ VE and Construct Reliability/CR often used to measure the validity of the model. A good validity of indicators would provide a value of VE ≥ 0.40 and value of CR ≥ 0.70 [17].

3.2. Assessment of structural model

SEM can examine the structural relationships between latent constructs. In evaluating the structural model, the researcher must focus on the relationship of interest to determine whether the hypothesized relationship in conceptualization phase is strongly supported by the data. The structural model requires two procedures for measurement. They are the fitness of model and evaluation of causal relationship with path analysis [29]. The model fit applies the same indexes in evaluating the measurement model (please refer to the fit indexes in Measurement Model). The second step is to measure the model of causal relationship. To simplify the study without violating the statistical significance, LISREL can reduce the latent variable 2nd order (indicators) by using latent variable score. The statistical test for the causal relation of the structural model with a significance level of 5% so that the critical value of t-value is ± 1.96. The estimation results of all the causal relationships are categorized into two; first, the output showing the (t-values) to assess if the relationship happens and second; the output shows the "standardized solutions" values to measure the degree of relationship between variables.

3.3. Data cleaning

It is important to acknowledge the importance of the data cleaning in SEM [30,31]. The data sample size is the first consideration. However, there is no agreement as regards the number of sample size needed in SEM. For example, Kline (2011) recommends that a sample size of 10 to 20 respondents per estimated parameter is sufficient. However, Kline (2011) describes that a sample size of fewer than 100 households is small sample size, the sample size between 100 and 200 households is medium sample size, and sample size that is greater than 200 are considered as large sample size [3]. While, Weston and Gore-Jr (2006) contends that 200 respondents are satisfactory when researcher do not face with technical difficulties like a missing data [32]. Secondly, Multicollinearity is another issue in data cleaning. Multicollinearity occurs when extremely observed variables are redundant. It is also required for studies to scrutinize univariate and multivariate outlier's factors. The respondent's response will characterize a univariate outlier when the responses are out of range on only one variable [33].

Multivariate outliers will exist if respondents have two or more extreme responses. Removing of multivariate outliers could solve the problem of multivariate outliers. Multivariate distribution of statistics must be normal to reduce the effect the fallacy of statistical tests that can influence the fitness of model. The analysis of skewness and kurtosis distribution of each observed variable aims to determine univariate normality. Transformation of data and deleting or transforming univariate or multivariate outliers enhances multivariate normality and increase data normality. Missing data signify a systematic loss of data, and it is a requirement to set up in data cleaning process. It is vital to acknowledge missing data before the study proceeds the data analysis through SEM. The treatment of missing data can use the the descriptive analysis tool in the SPSS application.

4. Built environment research using SEM

Particularly, in built environment studies, a set of indicators of questionnaire examines latent constructs. The first generation of statistical technique could not proceed latent constructs, and this is the

fundamental reason to use of SEM that allows the relationship among the constructs to be modeled with their respective item variables and for simultaneous analysis. SEM is a mix of factor analysis and path analysis to confirm a summary of the interrelationships among variables [32]. Researching in built environment field are always complicated situation and embedded multidimensional that need complex research questions through hypotheses approach. The first statistical generation to analyze built environment research could not deal with the complex tasks. The reason is it cannot easily allow for measuring and testing of multivariate models with latent variables in the one process.

In this sense, SEM allows for the testing of such models. SEM can provide a summary of variables, the hypothesized relationships, the constructs. The ability of SEM in estimating and testing the relationships among constructs is the advantage over the first development of generation statistical analysis method. The use of SEM application facilitates built environmental studies to conduct of numerous different multiple regression models and modifying through identification and removal of the weaknesses in the model until the model is fit to the data. The application of measures in SEM to represent constructs provides the establishment of the construct validity of factors unlike in general linear models where constructs may only have one measure representation. SEM involves measurement errors while these errors are not calculated in the general linear models in the first development generation statistical method.

5. Benefits and limitation of SEM application

According to Byrne (2010), the employment of SEM to advance the built environment to compares other statistical multivariate techniques are [17]:

1. SEM takes a confirmatory approach to data analysis by specifying the associations between variables. Other is descriptive by nature like exploratory factor analysis so that hypothesis testing is rather difficult to do.
2. SEM allows estimating of error variance parameters while others cannot assess or correct for errors. For instance, a regression analysis disregards the potential error in all the independent variables in a model, and this can result in the possibility of incorrect conclusions due to misleading regression estimation.
3. SEM combines unobserved (latent) and observed variables altogether yet others based on observed measurements only.
4. SEM can develop a model of multivariate relations, and estimate direct and indirect effects of variables under study while other techniques cannot perform such the task.

Despite the advantage of the SEM, there is various issue need to be considered. SEM demands large sample size to calculate the parameter of variances, regression coefficients and covariances based on Maximum Likelihood (ML), and to meet the normal distribution of the variables. The model based on a small sample size cause estimation problems and unreliable results. Built environment research to apply SEM requires the minimum 100 sample sizes to meet the assumption of maximum likelihood estimation [3]. The process of SEM is also technically complicated that may make studies to misuse the technique in developing a "fit index tunnel vision" [3]. Consideration of multiple fit indices and residuals can be ignored by studies in the testing fitness to the data but only consider a single index like CFI thereby avoid modification of the model.

6. Conclusion

In conclusion, SEM can provide advantages in built environment studies by considering more complex research questions and test multivariate in a single study. The employment of SEM includes the interaction of statistical procedures and theoretical understanding in built environment research. In spite of various benefits of the SEM, the paper highlighted some of the disadvantage issues. This study aims to deliver a better acknowledgment for using of SEM in the built environment study. It is though also recommended that researchers in built environment should be motivated to make more consultation to some of the references for more understanding of the SEM application. This paper is limited to the brief

introduction to the background, features of the SEM and its employment in built environment research. Finally, further study is recommended to examine the methodological approach to facilitate analysis in built environment research through SEM with case studies

References

[1] Carvalho J D and Chima F O 2014 *American International Journal of Contemporary Research* **4** 6
[2] Hoyle R H *Handbook of Structural Equation Modelling* (New York, London: The Guilford Press: A Division of Guilford Publications, Inc., 2012)
[3] Kline R B *Principles and Practice of Structural Equation Modeling*, ed D A Kenny and T D Little (New York, London: THE GUILFORD PRESS: A Division of Guilford Publications, Inc., 2011)
[4] Timothy T, Tsai L T and Yang C C *Applying Structural Equation Modeling (SEM) in Educational Research: An Introduction*, ed M S Khine (Rotterdam, Boston, Taipei: Sense Publishers, 2013)
[5] Awang Z *A Handbook on Structural Equation Modeling* (Malaysia: MPWS Rich Resources, 2014)
[6] Baumgartner H and Homburg C 1996 *International Journal of Research in Marketing* **13** 139
[7] Choi J N 2011 *Creativity Research Journal* **16** 187
[8] Nair H P, Kumar D and Ramalu S S 2015 *Asian Social Science* **11** 200
[9] Arbuckle J L *Amos 22 User Guide: Amos* (New York: Development Corporation, 2013)
[10] Bian H *Structural Equation Modelling with AMOS II* (New York: Routledge, 2011)
[11] A P Nair, H Kumar D and Sri Ramalu, S. 2014 *Asian Social Science,* **10** 145
[12] Dyer P, Gursoy D, Sharma B and Carter J 2007 *Tourism Management* **28** 409
[13] Manafi M and Subramaniam I D 2015 *Asian Social Science* **11** 131
[14] Bollen K A *Structural Equations with latent Variables* (New York: Willey, 1989).
[15] Daniel L O and Albert S 2013 *Structural Equation Modeling: A Multidisciplinary Journal* **20** 409
[16] Hair J F, Black W C, Babin B J and Anderson R E *Multivariate Data Analysis: Overview of Multivariate Methods* (New Jersey: Pearson Education International, 2010)
[17] Byrne B M *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming* (New York: Taylor and Francis Group, 2010)
[18] Taehun L and Robert C 2014 *Structural Equation Modeling: A Multidisciplinary Journal* **22** 102
[19] Tenko R and George A M 2016 *Structural Equation Modeling: A Multidisciplinary Journal* **23** 845
[20] Hu L and Bentler P M 1999 *Structural Equation Modeling: A Multidisciplinary Journal* **6** 1
[21] Jenny C, Victoria S & Carl F 2015 *Structural Equation Modeling: A Multidisciplinary Journal* **22** 517
[22] Barrett P 2007 *Personality and Individual Differences* **42** 815
[23] MacCallum R C, Browne M W and Sugawara H M 1996 *Psychological Methods* **1** 130
[24] Tabachnick B G and Fidell L S *Using multivariate statistics 2nd ed.* (New York: Harper & Row, 1989)
[25] Shevlin M and Miles J N V 1998 *Personality and Individual Differences* **25** 85
[26] Bentler P M and Chou C P 1987 *Sociological Methods & Research* **16** 78
[27] Milica M, David P and Roy L *Structural Equation Modeling: A Multidisciplinary Journal* **24** 666
[28] Mark H C and Myeongsun Y 2014 *Structural Equation Modeling: A Multidisciplinary Journal* **22** 236
[29] Bengt M and Tihomir A 2014 *Structural Equation Modeling: A Multidisciplinary Journal* **22** 12
[30] Felix T and Karthika M 2015 *Structural Equation Modeling: A Multidisciplinary* **22** 631
[31] Xin T, Zhiyong Z and Ke-Hai Y 2014 *Structural Equation Modeling: A Multidisciplinary Journal* **21** 553
[32] Weston R and Gore-Jr P A 2006 *The Counseling Psychologist* **34** 719

[33] Alberto M O, Amanda J F and Alexander G H 2017 *Structural Equation Modeling: A Multidisciplinary Journal* **24** 495