UNIVERSITAS WIRARAJA
LEMBAGA PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT

SURAT PERNYATAAN
Nomor : 068/SP.HCP/LPPM/UNIJA/II/2021

Yang bertanda tangan di bawah ini   :

Nama	: Nailiy Huzaimah, S. Kep., Ns, M.Kep.
Jabatan	: Sekretaris LPPM
Instansi	: Universitas Wiraraja

Menyatakan bahwa   :

1. Nama   : Dr. Anik Anekawati, S.Si., S.Pd., M.Si.
   Jabatan   : Staf Pengajar Fakultas Keguruan dan Ilmu Pendidikan

Telah melakukan cek plagiarisme ke LPPM menggunakan *software turnitin.com* untuk artikel dengan judul *"STRUCTURAL EQUATION MODELLING WITH THREE SCHEMES ESTIMATION OF SCORE FACTORS ON PARTIAL LEAST SQUARE (CASE STUDY: THE QUALITY OF EDUCATION LEVEL SMA/MA IN SUMENEP REGENCY"* dan mendapatkan hasil similarity sebesar 26%

Demikian surat pernyataan ini dibuat untuk digunakan dengan sebaik-baiknya.

Sumenep, 26 Februari 2021
an Sekretaris LPPM
Universitas Wiraraja,

Nailiy Huzaimah, S. Kep., Ns, M.Kep.
NIDN. 0727069003

Structural Equation Modelling with Three Schemes Estimation of Score Factors on Partial Least Square (Case Study: The Quality Of Education Level SMA/MA in Sumenep Regency)

by Anik Anekawati 260221

PAPER · OPEN ACCESS

Structural Equation Modelling with Three Schemes Estimation of Score Factors on Partial Least Square (Case Study: The Quality Of Education Level SMA/MA in Sumenep Regency)

View the article online for updates and enhancements.

Structural Equation Modelling with Three Schemes Estimation of Score Factors on Partial Least Square (Case Study: The Quality Of Education Level SMA/MA in Sumenep Regency)

[1]Anik Anekawati, [2]Bambang Widjanarko Otok, [3]Purhadi, [4]Sutikno

[1]Student of Doctoral Programme, Departement of Statistics, Institut Teknologi Sepuluh Nopember
[2,3,4]Departement of Statistics, Institut Teknologi Sepuluh Nopember

[1]anekawa10@gmail.com, [2] bambang_wo@statistika.its.ac.id,
[3]purhadi@statistika.its.ac.id, [4]sutikno@statistika.its.ac.id

Abstract.

Research in education often involves a latent variable. Statistical analysis technique that has the ability to analyze the pattern of relationship among latent variables as well as between latent variables and their indicators is Structural Equation Modeling (SEM). SEM partial least square (PLS) was developed as an alternative if these conditions are met: the theory that underlying the design of the model is weak, does not assume a certain scale measurement, the sample size should not be large and the data does not have the multivariate normal distribution. The purpose of this paper is to compare the results of modeling of the educational quality in high school level (SMA/MA) in Sumenep Regency with structural equation modeling approach partial least square with three schemes estimation of score factors.This paper is a result of explanatory research using secondary data from Sumenep Education Department and Badan Pusat Statistik (BPS) Sumenep which was data of Sumenep in the Figures and the District of Sumenep in the Figures for the year 2015. The unit of observation in this study were districts in Sumenep that consists of 18 districts on the mainland and 9 districts in the islands. There were two endogenous variables and one exogenous variable. Endogenous variables are the quality of education level of SMA/MA (Y1) and school infrastructure (Y2), whereas exogenous variable is socio-economic condition (X1). In this study, There is one improved model which represented by model from path scheme because this model is a consistent , all of its indicators are valid and its the value of R-square increased which is: $Y_1=0.651Y_2$. In this model, the quality of education influenced only by the school infrastructure (0.651). The socio-economic condition did not affect neither the school infrastructure nor the quality of education. If the school infrastructure increased 1 point, then the quality of education increased 0.651 point. The quality of education had an R^2 of 0.418, which indicates that 41.8 percent of variance in the quality of education is explained by the school infrastructure, the remaining 58.2% is explained by the other factors which were not investigated in this work.

Key Words: partial least square, schemes estimation of score factors

1. Introduction

Research in education often involves latent variables. Statistic analysis technique that can be used to analyse the relationship among the latent variables (LVs) as well as between latent variable (LV) and its indicators is Structural Equation Modeling (SEM). Schumacker and Lomax concluded in his research that SEM is often described using factors that are seen as LV [7]. Variable to be analysed is the latent variable which cannot be observed directly but can be observed using measured indicators.

Several estimation methods have been developed in SEM. The very famoust one is the Maximum Likelihood Estimation (MLE). Since the covariance-based SEM is based on large sample size theory, it must fullfil the requirements of normal multivariate distribution, Bollen therefore suggested an estimation method which accepts an abnormality and is asymptotically efficient [2]. This method is called Weighted Least Square (WLS). WLS is an estimation method adapted from Asymptotically Distribution Free (ADF) method which is the most common method due to its independence on the type of data distribution. Although WLS has advantages compared to MLE, but it requires a larger sample size. If the sample size is too small and not normal multivariate distribution, the researchers will face analysis results which are improper solution [1][3]. As an alternative solution of this problem, variant-based SEM was developed or also known as Partial Least Square (PLS).

SEM PLS is a stastistic technique that generalizes and combines factor analisys, Principal Component Analysis (PCA) and regression analysis by separate estimation procedure between latent variables and its indicators. Additionally, SEM PLS was developed as an alternative method when the theory underlying model design is weak, not based on many assumptions, has a small sample size, and does not have to be a normal multivariate distribution [6].

The shortcoming of SEM PLS is the use of unknown data distribution so that a value of model significance cannot be obtained and therefore a resampling approximation can be used. One of the resampling methods used for a small sample size is bootstrap since a small sample size tends not to distribute normally resulting in an inaccurate parametric statistic estimation. Bootstrap method works using resampling with replacement by creation of pseudodata using information of initial data including its characteristics so that a correct estimation can be achieved [5].

Main stage in PLS algorithm is an iteration process to generate a latent variable score through weight estimation. There are three main schemes in this process, namely centroid, factor and path scheme. In practice, the selection of one weighting scheme has a little relationship in estimation process [9]. Tenenhaus observed and mentioned that the three schemes do not influence the results significantly [8]. However, in the theoretical level, the schemes are very important to understand how PLS can be applied for a different technique from multiple table analysis.

The case of education quality improvement in Sumenep regency has a unique characteristic since it is divided in mainland and islands with a high disparity in education quality. Within the frame of equalization of education quality, a study is required which includes latent variable, namely the quality of the education in high school level, the school infrastructure and the socio-economic condition in Sumenep. Since the sample units are 27 districts, which falls under a small sample size category, it requires analysis technique that includes latent variables with a small sample size, namely SEM PLS.

Based on the above explanation, a research was conducted with the aim to compare modelling results of education quality of high school level at Sumenep regency using approximation method Structural Equation Modeling Partial Least Square with three schemes estimation of score factor.

2. Literature Review

Structural Equation Modeling with Partial Least Square (PLS) Approximation

SEM based on variance or Partial Least Square (PLS) consists generally of two sub-models which are measurement model or outer model and structural model or inner model. Outer model shows how observed variables represent latent variables to be measured, whereas inner model shows the estimation strength among latent variables.

Inner model in PLS describes the relationship among latent variables. The model equation can be written as the following [9]: $\eta = \beta_0 + \beta\eta + \Gamma\xi + \zeta$

Since PLS is designed for recursive model, in the latent variable relationship, each dependent latent variable **η** is called as causal chain system from latent variable and can be specified as:

$$\eta_j = \sum_i \beta_{ji}\eta_i + \sum_b \gamma_{jb}\xi_b + \zeta_j$$

Where β_{ji} and γ_{jb} are the path coefficient that connect LV endogenous (η) with LV endogenous and LV endogenous with LV exogenous (ξ) in the index range i and b, and where ζ_j is residual from the inner model with the value $E(\zeta_j) = E(\xi_i\zeta_j) = 0$.

Outer model is defined how each indicator block relates to its latent variable. Trujillo reported the outer model equation for block indicators of reflective model as an equation of simple regression [9]: $x_{jk} = \lambda_{jk}\xi_j + \varepsilon_{jk}$

where λ_{jk} is coefficient of loading factor from the relationship between the j-th LV (ξ_j) with k-th its indicator (x_{jk}), and ε_{jk} is error of each measurement variable.

Parameter estimation on SEM PLS model has three categories: weight estimate, path estimate and also mean and location parameter. Weight estimate is used to generate score of LV. Path estimate reflects the relationship between LVs and loading estimate between LV and its indicators. The last category relates to mean and location parameter for indicator and LV. In order to obtain these three estimations, PLS uses iteration proccess.

Weight Estimate

In weight estimation, there are three iteration procedures of simple regression or multiple regression considering the relationship of the inner model, the outer model and weight relation. The results of iteration procedures, that are a set of stabil weights, is used to calcualte the variable laten score which is a linear combination of their indicators [8].

Weight calculation is used to estimate LV score with the following steps:

1. Outer model estimation (outside approximation)

In the outside approximation, iterative process begins with an early initialization of each LV as a linear combination (or weighted aggregate) of indicators and written in equation: $l_j = \sum_k \tilde{w}_{jk}x_{jk}$

where $\tilde{w}_{jk}$ is weighted in outer model (outer weights). The basic idea of the outside approximation is to obtain a set of weights to estimate a latent variable accounting for as much variance as possible for the indicators and the constructs [9]. Chin suggested to set initial weights by equal value to present the first approximation of the latent variable as a simple sum of its indicators [9].

2. Inner model estimation (inside approximation)

In the inside approximation, iterative process takes into account the connections among LVs in the inner model to obtain initialization of each LV calculated as a weighted aggregate of its adjacent LVs. Definition of inside approximation is as follows $Z_j = \sum_{i:\beta_{i\leftrightarrow j}\beta_{j\leftrightarrow i}} e_{ji}l_i$

where e_{ij} is inner weights which can be choosen from three scheme which are centroid, factor or path. Each scheme is defined by Trujillo as follows [9] :

i. Inner weight (e_{ij}) in centroid scheme is equal to the sign correlation between l_i and l_j, where centroid scheme is defined as:

$$e_{ij} = \begin{cases} sign\{cor(l_j, l_i)\} & \xi_j\xi_i \ \text{adjacent} \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

ii. Inner weight (e_{ij}) in factor scheme is the correlation between l_i and l_j which considers not only the sign direction but also the strength of the paths in inner model. Factor scheme is defined as:

$$e_{ij} = \begin{cases} cor(l_j, l_i) & \xi_j \xi_i \text{ adjacent} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

iii. Inner weight (e_{ij}) in path scheme is the weighting of neighboring LVs depending whether of not the neighboring variables are antecedent or consecuent from LV which is to be estimated. Path scheme is defined as:

$$\left. \begin{aligned} e_{ij} &= cor(l_j, l_i); \text{if } \xi_j \text{is explained by } \xi_i \\ l_j &= \sum_i e_{ji} l_i; \; e_{ij} \text{ is coefficient in the regression of } l_i \text{ on } l_j \end{aligned} \right\} \quad (3)$$

3. Updating outer weight

Once the inside approximation is done, the internal estimation Z_j is considered by observing their indicators. This can be done by updating the weight in outer model (outer weight). The weight outer w_{jk} in the reflective model is expressed in the following equation: $w_{jk} = \left(Z_j' Z_j \right) Z_j' x_{jk}$

4. Checking convergence

In each iterative procedure such as S=1,2,3,..., convergence is checked by comparing outer weight in the S step over outer weight in the S-1 step. Wold suggested $\left| \tilde{w}_{jk}^{S-1} - \tilde{w}_{jk}^{S} \right| < 10^{-5}$ as a convergence criterion [9].

Path Estimate

Next step is the calculation of the path coefficient estimation and loading, $\hat{\beta}_{ji}$ and $\hat{\lambda}_{jk}$, in the inner and outer model. In the inner model, path coefficient is estimated using Ordinary Least Square (OLS) as the multiple linier regression analysis from relation between l_j and l_i, that is $l_j = \sum_i \hat{\beta}_{ji} l_i$ and $\hat{\beta}_{ji} = \left(l_i' l_i \right)^{-1} l_i' l_j$.

In the outer reflective model, loading coefficient is estimated as a simple linier regression from relation between x_{jk} and l_j, that is: $x_{jk} = \hat{\lambda}_{jk} l_j$ and $\hat{\lambda}_{jk} = \left(l_j' l_j \right)^{-1} l_j' x_{jk}$

Mean and Location Parameter Estimate

There are two locations parameters to be estimated, namely β_{0j} (constants in the inner model) and λ_{0jk} (constants in the outer reflective model). The equation with predictor specification above can be written as follows [9]: $E\left(\xi_j | \xi_i \right) = \beta_{0j} + \sum_i \beta_{ji} \xi_i$ for inner model and $E\left(x_{jk} | \xi_j \right) = \lambda_{0jk} + \lambda_{jk} \xi_j$ for outer reflective model.

These location parameter takes the mean of indicator and latent variables into account. Prior to the calculation of the location parameter, the mean to estimate the latent variables is defined as follows [8] : $\hat{m}_j = \sum_k \tilde{w}_{jk} \bar{x}_{jk}$ and $\hat{\xi}_j = l_j + \hat{m}_j$, so that the estimation of the location parameter β_{0j} (constant in the inner model) and λ_{0jk} (constant in the outer reflective model) can be interpreted as [9]:

$$\hat{\beta}_{0j} = b_{0j} = \hat{m}_j - \sum_i b_{ji} \hat{m}_i, \; \hat{\lambda}_{0jk} = \bar{x}_{jk} - \hat{\lambda}_{jk} \hat{m}_j \text{ and } \hat{\pi}_{0j} = \hat{m}_j - \sum_k \hat{\pi}_{jk} \bar{x}_{jk}$$

Evaluation of SEM PLS Model

The outer models with reflective indicator can be evaluated through convergent validity, discriminant validity and composite reliability. Inner model is evaluated by observing the percentage of explained variance with the value of R^2 for the dependent latent constructs using Stone-Geisser Q test [6]. Here is a more detailed description related to the evaluation of outer and inner model.

1. Outer model

Evaluation of outer models is an assessment of the reliability and validity of the variables or is defined how each block of indicator associated with LVs. There are three criteria for evaluating the measurement model, namely: convergent validity, discriminant validity, and composite realibility.

a. Convergent validity

Convergent validity of the measurement model relates to the principle that the indicators of a construct should be highly correlated and can be seen from the value of the loading factor. Assessment is based on the correlation between the item score/component score and construct score, which is calculated by PLS. The model is declared good if the correlation coefficient is 0.70 more than the construct to be measured. However, for the early stage of a research, scale measurement development with loading value of 0.5 to 0.6 is considered to be sufficient.

b. Discriminant validity

Discriminant validity of the measurement model relates to the principle that indicators of the different variable should be highly correlated and can be seen from value of cross loading. If the correlation between construct and the measurement item is larger than the other constructs, then it indicates that the latent construct predicts that the size of their block is better than the size of the other blocks.

Another method for seeing the discriminant validity is to compare the value of square root of extracted (AVE) of each construct with a correlation among the other constructs in the model. It has a good discriminant validity if the value of the square root of AVE of each construct is greater than the value of correlation between the construct and other constructs in the model [6]. Ghozali suggested value of AVE greater than 0.5[6].

c. Composite realibility

Composite realibility of indicator block that measures a construct can be evaluated by internal consistency and Cronbach's alpha. By using the outputs of PLS, the composite realibility can be calculated with the following formula:

$$\rho c = \left(\sum_{i=1}^{k} \lambda_i^2 \right)^2 \rho c \Bigg/ \left(\left(\sum_{i=1}^{k} \lambda_i^2 \right)^2 + \sum_{i=1}^{k} \mathrm{var}(\varepsilon_i) \right)$$

where λ_i is component loading of indicator and $\mathrm{var}(\varepsilon_i) = 1 - \lambda_i^2$.

2. Inner model (structural model)

Structural model is evaluated using R^2 to see the relationship among the LVs of the research model. Stone-Geisser Q-square test for predictive relevance and stability of these estimation results are evaluated by t-test statistics through bootstrapping procedure. The assessment criteria through PLS model as follows:

$$f^2 = \frac{R_{include}^2 - R_{exclude}^2}{1 - R_{include}^2}$$

where $R_{include}^2$ and $R_{exclude}^2$ is R-square from dependen LV when LV predictor is used or is taken out in model.

3. Research Methodology

This work is an explanatory research since the author did not possess initial data or an description of the education quality model. The data used is a secondary data from Sumenep Education Department which is a summary of data on education in SMA/MA in year 2015, Badan Pusat Statistik (BPS) Sumenep that is data of district's sumenep in figure for 2015 and Sumenep in figure for 2015.

The observation units in this research were the districts in Sumenep which consists of 18 districts on mainland and 9 districts on islands so that the total observation units were 27 districts. Additionally, there were 2 endogenous variables, namely the quality of education in SMA/MA (Y1) and the school infrastructure (Y2) and 1 exogenous variable, namely the socio-economic condition in Sumenep.

Indicators of the quality of education SMA/MA (Y1) are a ratio of crude participant figure SMA/MA (Y_{11}), the ratio of SMA/MA accredited at least B (Y_{12}), a ratio average of national exam score SMA/MA more than 7 (Y_{13}), a ratio SMA/MA that have ISO 9001:2008 certificate (Y_{14}).

Indicators of the school infrastructure (Y2) are ratio of student number in a class (Y_{21}), the ratio of minimum class size (Y_{22}), the ratio of laboratorium possession (Y_{23}), and the ratio of library possession (Y_{24}).

Indicators of socio-economic condition in Sumenep (X_1) are the ratio of the number of prosperous families (X_{11}), the ratio of household industries (X_{12}), the ratio of households using clean water (X_{13}), the ratio of households using basic sanitation (X_{14}), the ratio of the number of medical personnel (X_{15}), and the ratio of healthy house (X_{16}).

The analysis step in this research is as follows:

1. Conduct descriptive statistics as an initial overview on the education quality of high school in Sumenep regency in 2015 and all factors that are expected to affect them.
2. Build a conceptual model based on theory and construct path diagram.
   Determination of education quality model in high school level related to many dimensions such as infrastructure and facilities, and socio-economic condition in that region. From that conceptual model, a structural model is built with a hyphothesis that will be tested as follows:
   H_A : Socio-economic condition (X_1) and the school infrastructure (Y_2) having an influence on the quality of education in SMA/MA (Y_1).
   H_B : Socio-economic condition (X_1) having an influence on school infrastructure (Y_2).
3. Carry out validitation test on outer model
   Evaluation of each indicator variables can be done by observing the value of convergent validity, discriminant validity and the value of composite reliability by means of software Smart PLS.
4. Obtain score of latent variable from each of latent variable
   Estimation of latent variable score SEM model with using the PLS approximation. The score of an exogenous variable is $\hat{\xi}_j = l_j$, and the score of an endogenous variable is $\hat{\eta}_j = k_j$. Since the process to obtain the score of latent variable for both endogenous and exogenous is the same, the scores of latent variable are represented by l_j, as illustrated in Figure 1:
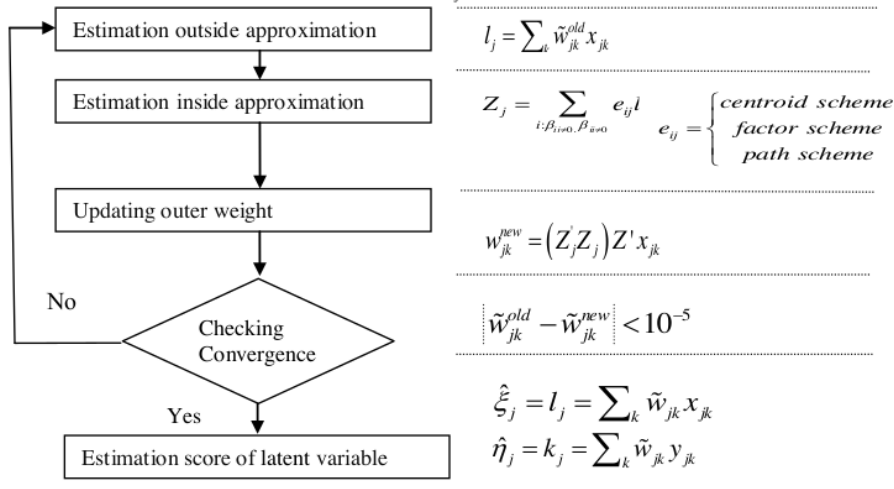

$$l_j = \sum_k \tilde{w}_{jk}^{old} x_{jk}$$

$$Z_j = \sum_{i:\beta_{ij\neq0}.\beta_{ii\neq0}} e_{ij} l \qquad e_{ij} = \begin{cases} centroid\ scheme \\ factor\ scheme \\ path\ scheme \end{cases}$$

$$w_{jk}^{new} = \left(Z_j'Z_j\right)Z' x_{jk}$$

$$\left|\tilde{w}_{jk}^{old} - \tilde{w}_{jk}^{new}\right| < 10^{-5}$$

$$\hat{\xi}_j = l_j = \sum_k \tilde{w}_{jk} x_{jk}$$
$$\hat{\eta}_j = k_j = \sum_k \tilde{w}_{jk} y_{jk}$$

Figure 1. Procedure to estimate the score of latent variable

5. Select the best model, interprete and conclude results.

4. Main Results
A. Evaluation outer model (measurement model)
a. Convergent validity

The loading factor of each indicator based on three weighted schemes, namely the centroid, factor and path is summarized in the following description. There are four indicators that have a loading factor less than 0.5 for all three schemes, namely X_{12}, X_{14}, X_{15} and X_{16}. It can be concluded

that the ratio of household industries, the ratio of households using basic sanitation, the ratio of the number of medical personnel and the ratio of healthy house are invalid or could not be used as an indicator to build construct of socio-economic condition.

The following is a summary of loading factor value of the school infrastructure for 3 schemes. All of the indicators have a loading factor more than 0.5 for three schemes. It can be concluded that the ratio of student number in a class, the ratio of minimum class size, the ratio of laboratorium possession, and the ratio of library possession are valid or could be used as an indicator to build construct of the school infrastructure.

The following is a summary of loading factor value of the quality of education in SMA/MA for 3 schemes. There is one indicator that has a value of loading factor less than 0.5, namely Y_{13}. It can be concluded that a ratio average of national exam score SMA/MA more than 7 is invalid or could not be used as an indicator to build construct of the quality of education in SMA/MA.

b. Discriminant validity

The following is a cross loading of each indicator based on three weighted schemes. A value of cross loading and $\sqrt{AVE}$ based on output software Smart PLS. The correlation between socio-economic condition and its indicators is higher than the correlation between socio-economic condition and indicator of other variables, except X_{14}. It shows that the socio-economic condition predicts indicator within a group is better than the indicator in the other groups, except the ratio of households using basic sanitation.

The following is the comparison result of the value $\sqrt{AVE}$ of socio-economic condition and the correlation between socio-economic condition and the other variables through three weighted schemes. The value of $\sqrt{AVE}$ of socio-economic condition is smaller than the correlation between socio-economic condition and the other variables. It indicates that socio-economic condition has a low discriminant validity.

Table 1. The $\sqrt{AVE}$ Value and Correlation Between Socio-Economic Condition and The Other Variables

Centroid			Factor			Path		
Corellation X_1		$\sqrt{AVE}$	Corellation X_1		$\sqrt{AVE}$	Corellation X_1		$\sqrt{AVE}$
Y_2	Y_1		Y_2	Y_1		Y_2	Y_1	
0.59	0.478	0.441	0.598	0.466	0.441	0.6	0.459	0.436

The correlation between the quality of education and its indicators is smaller than the correlation between the quality of education and indicators of other variables. It shows that the quality of education can not predict inidicators within a group is better than the indicators in the other groups.

Table 2. The $\sqrt{AVE}$ Value and Correlation Between the Quality of Education and the Other Variables

Centroid			Factor			Path		
Corellation Y_1		$\sqrt{AVE}$	Corellation Y_1		$\sqrt{AVE}$	Corellation Y_1		$\sqrt{AVE}$
Y_2	X_1		Y_2	X_1		Y_2	X_1	
0.610	0.478	0.688	0.615	0.466	0.688	0.634	0.459	0.684

However, the value of $\sqrt{AVE}$ of the quality of education is greater than the value of the correlation between the quality of education and the others latent variables, as shown in Table 2. It concludes that the quality of education has a high discriminant validity.

The correlation between the school infrastructure and its indicators is smaller than the correlation between the school infrastructure and indicators of other variables. It shows that the school infrastructure cannot predict inidicators within group is better than the indicators in the other groups.

However, the value of $\sqrt{AVE}$ of the school infrastructure is greater than the value of the correlation between the school infrastructure and the others latent variables, can be seen in Table 3. It concludes the school infrastructure has high discriminant validity.

Table 3. The $\sqrt{AVE}$ Value and Correlation Between the School Infrastructure and the Other Variables

Centroid			Factor			Path		
Corellation Y_2		$\sqrt{AVE}$	Corellation Y_2		$\sqrt{AVE}$	Corellation Y_2		$\sqrt{AVE}$
Y_1	X_1		Y_1	X_1		Y_1	X_1	
0.610	0.590	0.734	0.615	0.598	0.734	0.634	0.600	0.739

c. Composite realibility

The school infrastructure and the quality of education have the value of composite reliability more than 0.7, but the socio-economic condition has less than 0.7. These results can be concluded that the school infrastructure and the quality of education are reliable or indicators of their variables are consistently measuring their constructs. However, the socio-economic condition is not reliable or their indicators are not consistently measuring the socio-economic condition.

B. Evaluation of Inner Model (Structural Model)

The evaluation of the inner model can be conducted by viewing the value of R-Square, in which the changes of the value of R-Square can be used to describe whether or not exogenous latent variables significantly effect on the endogenous latent variable. The following table is a summary of the value of R-Square from the three schemes.

Table 4. The Value of R-Square

Latent Variable	R-square		
	Centroid	Factor	Path
The school infrastructure	0.349	0.349	0.36
The quality of education	0.393	0.393	0.411

Based on Table 4, it can be concluded that the variance of the school infrastructure can be explained by the socio-economic condition amounted to 34.9% (using the centroid and factors scheme) and the remaining 65.1% is explained by other factors outside this research. The variance of the quality of education can be explained by socio-economic condition and the school infrastructure amounted to 39.3% (using the centroid and factors scheme) and the remaining 60.7% is explained by other factors outside this research. Both the school infrastructure model and the quality of education model are in the category of moderate models.

C. Significance Test

The significance test of the model cannot be done using In SEM PLS method, since the data distribution is unknown. Hence, the significance test was carried out using bootstrapping resampling method. The following are the results of significance tests both outer and inner models using $\alpha = 10\%$ with the value of the T-table of 1.65.

a. Ouer Model

The following is a result summary of the validity significance test of latent variables.There are two valid indicators that form the socio-economic condition, namely the ratio of the number of prosperous families and the ratio of households using clean water in the centroid and factor schemes. In contrast, there is only one valid indicator in the path schema, namely the ratio of the number of prosperous families. All of indicators are valid forming of the school infrastructure, except Y_{24} on the centroid scheme, i.e. the ratio of library possession. There is one indicator invalid forming of the

quality of education for three schemes, namely a ratio average of national exam score SMA/MA more than 7.

b.   Inner Model

The centroid and path scheme gave the same result, namely, the school infrastructure significantly affects the quality of education, but the socio-economic condition does not affect the school infrastructure and the quality of education. In the factor scheme, the socio-economic condition affects the school infrastructure, but does not affect the quality of education. While, the school infrastructure significantly affects the quality of education

D.   Model Improvements and Discussion

Model improvement was done by eliminating the invalid indicators either based on the results of convergent validity test and significance tests. Indicators that remain to be used in the socio-economic condition are X_{11}, X_{13} (for centroid and factor) and X_{11} (for path). Indicators that remain to be used in the quality of education are Y_{11}, Y_{12}, Y_{14} (for all of schemes). Indicators that remain to be used in the school infrastructure are Y_{21}, Y_{22}, Y_{23} (for centroid) and Y_{21}, Y_{22}, Y_{23}, Y_{24} (for factor and path).

After the model improvements, the evaluation is done both outer and inner models. The following are the results of the evaluation of the model has been improved.

The result of convergent validity test from three schemes shows valid, because all indicators of three latent variables have a loading factor above 0.5. The result of discriminant validity test shows that the correlation between latent variables and their indicators are higher than the correlation between latent variable and indicators of other variables for three latent variables and for the three schemes. Likewise, the $\sqrt{AVE}$ value of all latent variables is greater than the correlation between the latent variable and the others latent variables three schemes. The value of composite reliability for three latent variables of the three schemes are more than 0.7. It can be concluded that the indicators on each of the latent variables are valid and reliable. However, the result of significance test of outer model shows only indicator X_{11} on centroid scheme significanty invalid.

The following table shows the result of significance test of inner model for three schemes

Table 5. The Result of Significance Test of Improvement Model

Relation	Centroid		Factor		Path	
	T-value	Result	T-value	Result	T-value	Result
$X_1 \rightarrow Y_2$	3.859	affected	2.444	affected	1.258	not affected
$Y_2 \rightarrow Y_1$	3.407	affected	3.017	affected	6.848	affected
$X_1 \rightarrow Y_1$	0.270	not affected	0.287	not affected	0.153	not affected

The following is the evaluation of an improved model of the three schemes. Model of the centroid scheme is inconsistent and has one invalid indicator. Model of the factor scheme is a consistent , all of its indicators are valid but its the value of R-square decreased. Model of the path scheme is a consistent , all of its indicators are valid and its the value of R-square increased. So there is one improved model which represented by model from path scheme which is: $l_1 = 0.651 l_2$ and the full model as illustrated in Figure 2.

The model in the path scheme, the quality of education influenced only by the school infrastructure (0.648). The socio-economic condition did not affect neither the school infrastructure nor the quality of education. If the school infrastructure increased 1 point, then the quality of education increased 0.648 point. The quality of education had an R^2 of 0.418, which indicates that 41.8 percent of variance in the quality of education is explained by the school infrastructure, the remaining 58.2% is explained by the other factors which were not investigated in this work, such as support from local

or central government, sponsors, cooperation with the stakeholders of the school, the quality of teachers, the quality of the new students etc.
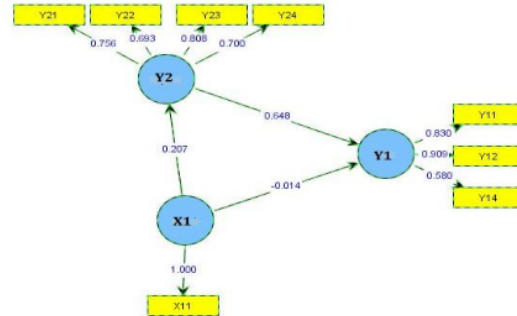

Figure 2. improved model from path scheme

The model in the path scheme shows that the socio-economic condition did not affect the quality of education. This condition contradicts the condition in the district sumenep since there are 185 schools only 13 of which have the status of public schools. It means that almost 93% of schools in the district Sumenep are private schools that were established through local community's initiative and the majority of them are religious schools. It reflects that the motivation of schools establishment tend to lead to idealism motivation, and less attention to the quality of education.

Trujillo stated that the selection of a certain weighted scheme than others has a few associated with the estimation process, whereas R-square of both models obtained in this work are not far different [9]. This agrees with observations made by Tenenhaus on the difference in the weighting results in the three schemes, in which no significant influence was shown [8]. The improved model which obtained in this was categorized as moderate models, which agrees with Chin's opinion [4].

5. Recommendations

It is necessary to test the effect of the spatial and necessary to add the variable of teachers' quality

References
[1]. Anderson JC and Gerbing DW 1984 The Effec of Sampling Error on Convergent, Improper Solution, and Goodnes of fit Indices for Maximum Likelihood Factor Analysis *Psychometrica* **49**.
[2]. Bollen K 1989 *Structural Equations With Latent Variables* (John Wiley & Sons, Inc, New York).
[3]. Boomsma A 1982 The *Robustnes of LISREL Against Small Sample Sizes in Facgtor Analysis Model*, In H. Wold and K. Joreskog (Eds), *System Under Inderec Observation* (New York Elsevier North Holand)
[4]. Chin WW 1998 *The partial least squares approach for structural equation modeling* Eds. Modern Methods for Business Research (Marcoulides GA)
[5]. Efron B 1979 Bootstrap Methods: another Look at the Jackknife *Annals of Statistics* **7** 1-26.
[6]. Ghozali I 2011 *Structural Equation Modelling Metode Alternatif dengan Partial Least Square PLS* (Undip Semarang)
[7]. Schumacker and Lomax 2004 *A Beginner's Guide To Structural Equation Modeling* (Lawrence Erlbaum Associates Inc New Jersey).
[8]. Tenenhaus M, Vinzi, Chatelin, and Lauro 2005 PLS Path-Modeling *Computational Statistics & Data Analysis* **48** 159–205.
[9]. Trujillo GS 2009 *PATHMOX Approach : Segmentation Trees in Partial Least Squares Path Modeling* (Universitat Politècnica de Catalunya)

Structural Equation Modelling with Three Schemes Estimation of Score Factors on Partial Least Square (Case Study: The Quality Of Education Level SMA/MA in Sumenep Regency)

Internet Source

1%

8   Jarot S. Suroso, R. Bambang Budhijana, Rahma Indah Delfiani. "The Effect Of Successful Enterprise Resource Planning (ERP) Systems On Employee Performance", 2018 International Conference on Orange Technologies (ICOT), 2018
Publication

1%

9   Submitted to Universitas Nasional
Student Paper

1%

10   www.emerald.com
Internet Source

1%

11   P R Satiawan. "The Sea is Not an Empty Space", IOP Conference Series: Earth and Environmental Science, 2017
Publication

1%

12   www.mdpi.com
Internet Source

1%

13   Ika Wirawati, Nur Iriawan, and Irhamah. "Bayesian Hierarchical Random Intercept Model Based on Three Parameter Gamma Distribution", Journal of Physics: Conference Series, 2017
Publication

<1%

Budi Yuniarto, Robert Kurniawan. "Modified

14 multiblock partial least squares path modeling algorithm with backpropagation neural networks approach", AIP Publishing, 2017
Publication

<1%

15 Noper Ardi, Isnayanti. "Structural Equation Modelling-Partial Least Square to Determine the Correlation of Factors Affecting Poverty in Indonesian Provinces", IOP Conference Series: Materials Science and Engineering, 2020
Publication

<1%

16 Dadang Syarif Sihabudin Sahid. "Learner Behavior in e-Learning as a Multicriteria Attribute based on Perspective of Flow Experience", International Journal of Advanced Computer Science and Applications, 2020
Publication

<1%

17 kyutech.repo.nii.ac.jp
Internet Source

<1%

18 journal.perbanas.ac.id
Internet Source

<1%

19 N F Rayesa, W A Gutama, Ferisna. "What affects customers to make online grocery purchasing?", IOP Conference Series: Earth and Environmental Science, 2020
Publication

<1%

20 Submitted to Universidad Estadual Paulista
Student Paper

<1%

21 Submitted to Universitas Brawijaya
    Student Paper   <1%

22 Simona Balzano, Laura Trinchera. "Chapter 4
    Structural Equation Models and Student
    Evaluation of Teaching: A PLS Path Modeling
    Study", Springer Science and Business Media
    LLC, 2011   <1%
    Publication

23 www.yumpu.com
    Internet Source   <1%

24 I Ralasari, R Setiadi, S M Soemartojo.
    "Relationships pattern of social media addiction
    level to morality level, apathetic level and
    academic score", Journal of Physics:
    Conference Series, 2020   <1%
    Publication

25 AFMM Rahman, J Solanky, MA Mrdha, AA
    Amrei. "Lowest Birth Weight Baby Survived in
    Abha General Hospital, Abha, K S A",
    Bangladesh Journal of Child Health, 2012   <1%
    Publication

26 download.atlantis-press.com
    Internet Source   <1%

27 Wallace Giovanni Rodrigues do Valle, Mariana
    Rodrigues de Almeida. "Chapter 97 Application
    of the Structural Equations' Modeling to Assess   <1%

Student Satisfaction", Springer Science and Business Media LLC, 2020
Publication

28 www.emeraldinsight.com
Internet Source   <1%

29 Submitted to De La Salle University - Manila
Student Paper   <1%

30 Submitted to University Tun Hussein Onn Malaysia
Student Paper   <1%

31 eprints.perbanas.ac.id
Internet Source   <1%

32 Nina Fei, Youlong Yang, Xuying Bai. "One Core Task of Interpretability in Machine Learning — Expansion of Structural Equation Modeling", International Journal of Pattern Recognition and Artificial Intelligence, 2019
Publication   <1%

Exclude quotes   On   Exclude matches   < 10 words
Exclude bibliography   On