Improving Accuracy of Educational Research Conclusions by Using Lisrel

Awaluddin Tjalla*
Department of Guidance and Counseling, State University of Jakarta, Kampus UNJ, Jl. Rawamangun Muka, Rawamangun, Jakarta
*Corresponding author: awaluddin.tjalla@yahoo.com

Received February 14, 2015; Revised April 26, 2015; Accepted April 28, 2015

Abstract The purpose of this paper are (1) to propose applying a technique of data analysis of educational variables using LISREL, (2) to construct variables including in the research model. The other benefit of this technique compared to the conventional analysis model are: (1) faulty estimate on variables relation caused by measurement error can be corrected and (2) statistical test on whether or not a theoretical model describing relation structure between variables can be carried out. The use of LISREL technique to analyze variable data for education is a must. Sharpness and accuracy in predicting the variables which are considered to have influence on variables can be obtained. On the contrary, measurement error which takes place from relation between research variables can be explained. Accordingly, this analysis technique is considered “comprehensive” to improve accuracy of conclusion generalization in the field of educational research which currently undergoes more complex problem.

Keywords: data analysis, educational variables, LISREL technique


1. Introduction

Globalization which is characterized by advancement in technology has resulted in problems on human life interaction, including those in educational field. In the perspective of research methodology, augmenting problems in educational field either quantitatively or qualitatively require settlement with correct technique of data analysis. This is aimed at obtaining objective result of data analysis and can be generalized accurately. Data analysis in educational field is not as easy as that in the other fields, such as : industry, engineering, agriculture, economics, etc., in which their required statistical analysis can easily be fulfilled. Interpretation made does not bear high risk compared to the more complex data of educational variables. Therefore, correct technique of statistical analysis is required.

Pophamand Sirotnik (1973) mentioned two main objectives of statistical use in a bid to analyze variables data in the educational field, they are [19]:

1. Statistical techniques to describe data (descriptive statistics). This statistics is used to infer numerical data, such as : test scores, age and educational year.
2. Statistics used by researchers to describe better inference against a phenomenon observed at sample and then generalized conclusion of population is taken. This analysis technique points to relation between variables. In such case, educational researchers try to describe relation between such variables as students’ IQ, achievement and attitude to learning program at school.

With regard to the second objective of using statistics (inferential statistics), the commonly used technique of statistical analysis in the educational field is regression analysis by applying production function approach (Draper and Smith, 1981) [3]. However, such model constitutes underlying weaknesses in terms of : (1) concept, (2) result measurement, and (3) some biased sources which are the characteristics in the educational field. Accordingly, by applying regression with Ordinary Least Squares (OLS) approach, such disturbances need serious attention and required assumption can be met. This is because assumption break may cause incorrect and misleading conclusion generalization due to possible misinterpretation (Suriasumantri, 2000) [22].

2. Analysis

Path analysis is commonly applied as well in analyzing educational variables data irrespective of its weaknesses. The following are the weaknesses (Mueller, 1996) [18]: (1) When assumption of series of variables events (one variable cannot precede the other and vice versa) is not met, then path analysis cannot be applied; (2) It is always assumed that error outside the system has no correlation between one and another; (3) Difficulty in determining unknown parameters of the available data; (4) Direction of cause based relation cannot be determined by analysis result but dependent upon the concept developed by researchers. In case of any faulty concept, then conclusion and interpretation will be faulty as well; and (5) Principally, path analysis is the application of equation of
the Two Stage LeastSquares (TSLS) which bases as well on the Ordinary Least Squares (OLS). Accordingly, when the required assumption by OLS is not fulfilled, then the result will be biased too.

In comparison with regression analysis, path analysis possesses some excellence (James, Mulaik, and Brett, 1982) [6]. This is because we can make influence decomposition of the variables, know how huge the direct and indirect influence as well as that which does not belong to the cause based one. Besides, we can arrange an interesting model on cause based relation between free and tied variables. However, difficulty frequently arises when it comes to making or arranging series of events of such variables.

![Figure 1. LISREL model consisting of measurement and structural components](image)

At the measurement model analysis, latent variable cannot be directly measured (as a factor or construct, such as children’s spacial ability). Loading at each observed variable for one factor shows correlation between the researched construct and at the common variance with other variable identified as latent variable (Kim and Mueller, 1978) [13].

Equation model used for component of measurement equation is as follows:

1. For measurement equation model x (exogenous model) is:

\[ X = \Lambda_x + \delta; \]
\[
\begin{bmatrix}
X_1 \\
\vdots \\
X_q
\end{bmatrix} =
\begin{bmatrix}
\xi_1 \\
\vdots \\
\xi_n
\end{bmatrix}
+ 
\begin{bmatrix}
\delta_1 \\
\vdots \\
\delta_q
\end{bmatrix}
\]

(\(p x 1\)) (\(p x m\)) (\(m x 1\)) (\(p x 1\))

2. For measurement equation model y (endogenous variable) is as follows:

\[ X = \Lambda_y \eta + \epsilon; \]
\[
\begin{bmatrix}
y_1 \\
\vdots \\
y_q
\end{bmatrix} =
\begin{bmatrix}
\eta_1 \\
\vdots \\
\eta_m
\end{bmatrix}
+ 
\begin{bmatrix}
\epsilon_1 \\
\vdots \\
\epsilon_q
\end{bmatrix}
\]

(\(p x 1\)) (\(p x m\)) (\(m x 1\)) (\(p x 1\))

Where:

- \(X\) = vector of measured variable, denoting indicator of latent exogenous variable, \(\xi\).
- \(\Lambda x\) = lambda \(x\) = loading factor of \(x\) variable at a ksi factor.
- \(\xi\) = ksi = vector of latent variable.
- \(\delta\) = delta = vector of unique component (measurement error).
- \(\eta\) = vector of measured variable, denoting indicator of latent endogenous variable, \(\eta\).
- \(\Lambda y\) = lambda \(y\) = loading factor of \(y\) variable at eta factor.
- \(\eta\) = eta = vector of latent variable.
- \(\epsilon\) = epsilon = vector of disturbance component (measurement error).

Both of the above measurement equations (for \(x\) and \(y\)) are the same, showing relation between measured variables \(x\) and \(y\) with their latent variables (factor to be measured) \(\xi\) and \(\eta\).

Structural equation model points to latent variables analysis (Loehlin, 1992) [15]. This term is used as structural equation model determines relation between latent variables. Notation used for structural equation model is as follows:

\[ \eta = B \eta \Gamma + \zeta \]

Where:

- \(\eta\) = vector of endogenous latent variable (effect)
- \(\xi\) = vector of exogenous latent variable (cause)
- \(\zeta\) = vector of residual variable
- \(\Gamma\) = coefficient matrix describing impact of exogenous variable(\(\xi\)) to endogenous variable(\(\eta\))
- \(B\) = coefficient matrix describing impact of endogenous variable(\(\eta\)) to endogenous variable(\(\eta\)).

Basically, structural equation model belongs to equation in the form of matrix of parameters, exogenous variables, endogenous variables and residual. Detail or element of each matrix is much dependent upon the number of latent variables used and line of cause based relation described. Hence, application of the structural equation highly depends on cause based relation model designed to describe a phenomenon.

Statistically, there are 7 (seven) assumptions to be fulfilled (Mueller, 1996) prior to conducting a test to structural model, they are [18]:

1. Exogenous and endogenous latent variables possess the same average as zero \(E(\zeta) = 0\);
2. Structural relation of exogenous to endogenous latent variables is linear;
3. Error for \(\zeta\) at structural equation model: (a) having average same as zero \(E(\zeta) = 0\); (b) independent, and (c) has no correlation with latent variable \(E(\zeta \zeta \zeta) = 0\);
4. Matrix \((I - B)\) is not single, a matrix which cannot be inversed.
5. Average of observed variable of the exogenous and endogenous variables is 0, that is \( E(X) = E(Y) = 0 \).
6. Relation between indicatoros of exogenous and endogenous variables and latent variables (either exogenous or endogenous) is linear.
7. Error measurement forand in measurement equations \( x \) and \( y \): (a) possesses average of 0 \( [E(\delta) = E(\varepsilon) = 0] \); (b) independent; (c) exogenous and endogenous latent variables are not correlated \([E(\xi\delta \prime) = E(\delta\varepsilon \prime) = 0; E(\eta\delta \prime) = E(\delta\eta \prime) = 0 ; E(\eta\varepsilon \prime) = E(\varepsilon\eta \prime) = 0]\); and (d) do not have any correlation between measurement error \([E(\varepsilon\delta \prime) = E(\delta\varepsilon \prime) = 0]\).

In LISREL model there are two stages which are related one another, they are: (1) to test the model truth by seeing whether there exists significant difference between model and data; and (2) in case of any accord between model and data, a test of hypothesis on structural relation in such model can be carried out (Hair et al., 1998) [4]. To test the fit between theoretical model proposed with data a goodness-of-fit test can be applied. This test describes how far the arranged theoritical model fits to describe data. In furtherance, the proposed LISREL model is exhibited at Figure 2.

Remarks:

- \( \xi \) = Laten Variable
- \( \varepsilon \) = Social economic status
- \( \eta \) = % of white-skinned at school
- \( \eta \) = Previous learning achievement
- \( \eta \) = Popularity among white-skinned children
- \( \eta \) = Current learning achievement
- \( \delta \) = Faulty measurement of variables equation\( \gamma \)
- \( \delta \) = Faulty measurement of variable \( x \)
- \( \gamma \) = Coefficient which describes the impact of exogenous variable \( \xi \) to endogeneous variable \( \eta \)
- \( \beta \) = Coefficient which describes the impact of endogeneous variable \( \eta \) to exogeneous variable \( \xi \).
- \( \phi \) = Observe Variable
- \( X \) = Social economic status
- \( Y \) = % of white-skinned at school
- \( Y \) = Intelligence
- \( Y \) = Class average score (5)
- \( Y \) = Popularity among white-skinned children
- \( Y \) = Class average score (6)
- \( Y \) = Reading achievement

In furtherance, by applying LISREL as an analysis tool proposed by Maruyama and Miller (1979) [17], it is obtained that the value \( \chi^2 \) is 4.88 and not significant at the significance level of 5% \((p > \alpha)\) with sample of 154 students. This result shows that the proposed LISREL model is fitted with the data. The coefficient magnitude between exogeneous and endogeneous variables as well as between endogeneous and other endogeneous variables is seen at Table 1.

![Figure 2. LISREL model on impact of class mixing to learning achievement](image)

In order to obtain a model which is in accord with data, chi-square \( \chi^2 \) achieved has to possess probability \( \geq 0.05 \) (Hayduk, 1988) [5]. If zero hypothesis which states that there is no distinction between model with data cannot be rejected (not significant), it means there is no difference between model and data. In other words, the proposed theoretical model fits to describe data. In furtherance, the fit test between model and data can also be carried out by seeing value magnitude of GFI, AGFI and RMR. The more huge the value of GFI and AGFI obtained (ranging from \( 0 \) – \( 1 \)) and the smaller the value of RMR, it means that the proposed model is good. However, a model test by an index other than \( \chi^2 \) is non-probabilistic (cannot be used to determine significance level of the existing difference).

The research is conducted by Maruyama and Miller (1979) [17], in USA by applying LISREL to analyze data from Lewis dan St. John (1974) on influence of class mixing to learning achievement[14]. Reference used is that norms and values possessed by white-skinned students from middle class and achievement oriented will be accepted and internalized by black-skinned students if the white-skinned students are friendly to them. The proposed LISREL model is exhibited at Figure 2.

Table 1. Result of Analysis of Structural Equation Model on Direct Impact of Exogenous Variable to Endogenous Variables as well as Endogenous Variable to Other Endogenous Variable

<table>
<thead>
<tr>
<th>No.</th>
<th>Parameter</th>
<th>Estimate Value</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>( \gamma_{11} )</td>
<td>.27</td>
<td>Significant</td>
</tr>
<tr>
<td>2.</td>
<td>( \gamma_{12} )</td>
<td>.29</td>
<td>Significant</td>
</tr>
<tr>
<td>3.</td>
<td>( \gamma_{21} )</td>
<td>.07</td>
<td>Not significant</td>
</tr>
<tr>
<td>4.</td>
<td>( \beta_{21} )</td>
<td>.38</td>
<td>Significant</td>
</tr>
<tr>
<td>5.</td>
<td>( \beta_{31} )</td>
<td>.98</td>
<td>Significant</td>
</tr>
<tr>
<td>6.</td>
<td>( \beta_{32} )</td>
<td>.02</td>
<td>Not significant</td>
</tr>
<tr>
<td>7.</td>
<td>( \phi_{12} )</td>
<td>.06</td>
<td>Not significant</td>
</tr>
</tbody>
</table>
Based on the analysis result, it is shown that parameters $\gamma_{11}$, $\gamma_{21}$, $\beta_{13}$, and $\beta_{14}$ are significant, while parameters $\gamma_{22}$ and $\beta_{23}$ are not significant. The criteria to determine whether it is or not significant through estimate value from value $t \geq 1.96$ (Byrne, 1998) [2].

The above analysis result shows that:
1. Family SES ($\xi_1$) has influence on the previous learning achievement ($\eta_1$);
2. Percent age (%) of white-skinned at school ($\xi_2$) has influence on the previous learning achievement ($\eta_1$) and does not influence the popularity of the white-skinned ($\eta_1$);
3. The previous learning achievement ($\eta_1$) has influence on the current learning achievement ($\eta_1$);
4. Popularity among white-skinned ($\eta_2$) has no influence on the current learning achievement;
5. There is no significant relation between family SES ($\xi_1$) and percentage (%) of white-skinned at school ($\xi_2$).

Based on the hypothesis proposed by Maruyama and Miller on transmission of cultural value[17], if correct, then acceptance of black-skinned children by white-skinned children has influence on learning achievement of black-skinned children positively and substantially. However, with insignificant coefficient value at $\beta_{12}$ (.02), showing that the hypothesis of cultural transmission is not correct and that the previous learning achievement influences the current one and popularity or acceptance by the white-skinned children.

3. Discussion

To overcome the weaknesses in applying analysis techniques, it is proposed to use a model which is currently developed and made popular in a bid to analyze educational variables data known as LISREL (Linear Structural Relationships). This analysis technique is also commonly called “analysis of covariance structures, the moments structure model, and latent variable equation systems in structured linear models” (Long, 1983) [16].

This analysis technique belongs to culmination of contemporary methodology (Kerlinger, 1986) [12]. This technique allows inclusion of latent construct variables, in which it cannot be included in regression model and path analysis. The excellence of this model compared to the conventional ones such as regression and path analyses are: (1) estimate error on relation between variables due to measurement error can be well corrected and (2) statistical test on whether a theoretical model can be accepted or not which describes relation structure between variables can be carried out.

LISREL analysis techniques to examine the fit between the theoretical model with data. In case of no significant distinction between covariant matrix expected based on model or theory and that obtained from data, it can be interpreted as proof of the model’s truth. When a theoretical model supported by data (empirical evidence) has been found, the meaningfulness of each coefficient which is based on cause relation that exists in the model can be tested as well (JoreskogdanSorbom, 1984) [8].

In LISREL technique, there are two groups of equations (Schumacker and Lomax, 1996) [21], they are measurement and structural equations. The former belongs to factor analysis/s used for explorative analysis and aimed at stipulating measured variables which can be made as a good indicator for latent variables. In other words, measurement model is made to estimate factor loading for each variable which is theoretically stipulated as an indicator of latent variables or factors. The latter is used to describe cause based relation between latent variables, either between causing variables (exogenous) or between effect variables (endogenous) as well as being used to describe the magnitude of explained and unexplained variances.

4. Conclusion

The use of LISREL in analyzing educational variables is a must as today this analysis model belongs to the most comprehensive one for analysis of behavioral variables. Sharpness and accuracy in predicting variables are considered to have influence on other variables can be executed. In addition, wrong measurement arising out of relation between variables can be explained. Accordingly, this analysis technique is considered “comprehensive” to improve accuracy of conclusion generalization in the field of educational research which currently undergoes more complex problem.

In order to apply LISREL program, supporting knowledge and skill are required as to facilitate us in using this technique. This includes knowledge of mathematics and computer skill besides theoretical mastery to the research model proposed by the researchers.

**References**


