# Structural Equation Modeling for WTC as L2

Leoncio P. Olobia*
leoncio.olobia@lnu.edu.ph
Leyte Normal University
6500 Tacloban City, Philippines

**Abstract**

This paper is designed to explain Willingness to Communicate (WTC) in English as a second language (L2). Identified predictor variables constitute familiarity, self-perceived communication competence (SPCC), and motivation that influence the variability of WTC as criterion variable.

What the modeling techniques signify are various scenarios where predictability of WTC as a criterion variable will be ascertained using pathway analysis in increasing levels – simple pathway analysis to determine prediction of WTC based on each predictor such as familiarity, SPCC and motivation and to a more structured relationship established among interaction of predictor variables in different arrow indications as they predict Willingness to Communicate (WTC), deemed important for the researcher. Likewise, simple linear regression to multiple linear regression analysis indicate how each predictor variable will predict variations on WTC.

Keywords: familiarity, motivation, SPCC, WTC, structural equation modeling (SEM).

## 1. Introduction

This paper is designed to explain Willingness to Communicate (WTC) in English as a second language (L2). Identified predictor variables constitute familiarity, self-perceived communication competence (SPCC), and motivation that influence the variability of WTC as criterion variable.

Before proceeding to Structural Equation Modeling (SEM), the following descriptions of modeling techniques such as path analysis, SEM and network analysis are presented and discussions on SEM proceeds thereafter.

### 1.1 Structural Equation Modeling

Structural equation modeling (SEM) is a multivariate statistical framework that is used to model complex relationships between directly and indirectly observed (latent) variables. SEM is a general framework that involves simultaneously solving systems of linear equations and encompasses other techniques such as regression, factor analysis, path analysis, and latent growth curve modeling (Stein, Morris & Nock, 2012).

In brief, structural equation model (SEM) is like multiple linear regression model that predicts the outcome of criterion variable (dependent variable) based on correlative relationship with given predictor variables (independent variable). For instance, determining willingness to communicate (WTC) in English as second language (L2) in a given context can be predicted by the interaction of predictor variables such as familiarity, self-perceived communication competence (SPCC), and motivation. In simple multiple linear regression analysis, the linear path provides an ideological correlative relationship as correlation coefficient determines such relationship. In addition to this, SEM extends structural and correlational analysis through the inclusion of mediating variables that explain causality between predictor and criterion variables. Mediating variables ask the question “why” such causality exists. In the given example, familiarity as influencing WTC can be due to the existence of how a person is related to the group he is communicating with that affects WTC in a given situation. This is a latent variable that goes beyond correlational or predictive relationship as causality is established. In another instance, SPCC as another predictor variable can be caused by second language mastery which can also influence familiarity, hence, WTC. In this case, causality includes language mastery aside from how a person relates to the group as predictor variables. Many mediating variables that explain causes will be a focus in SEM. The presence of mediating variable in this instance gives a more robust understanding of the nature of relationship between predictor and criterion variables.

1.2 Path Analysis

Similar to SEM in some ways, path analysis explains correlational relationship between exogenous and endogenous variables but measurement is not done. There are some criteria used in path analysis, namely: rectangles or squares are used to denote observed variables while circles are used for unobserved variables, triangles used for constant variables.

Path analysis is normally exhibited using path diagrams that includes boxes for observed variables and circles for latent or unobserved variables with arrows indicating direction. This is like simple linear or multiple linear regression in path diagrams instead of measuring correlational coefficient, for instance.

Multiple regression analysis and path analysis share similarities as both predict correlational relationship between predictor and path analysis. However, path analysis can be more advantageous in the sense that it can explain causality among predictor variables which multiple regression does not dwell on.

1.3 Network Analysis

In brief, network analysis can be explained between points A and B where getting from A to B requires a line in one scenario. In another scenario, getting from A to B may follow a path aside from a straight line. Network analysis is explained as “things” that travel along a path.

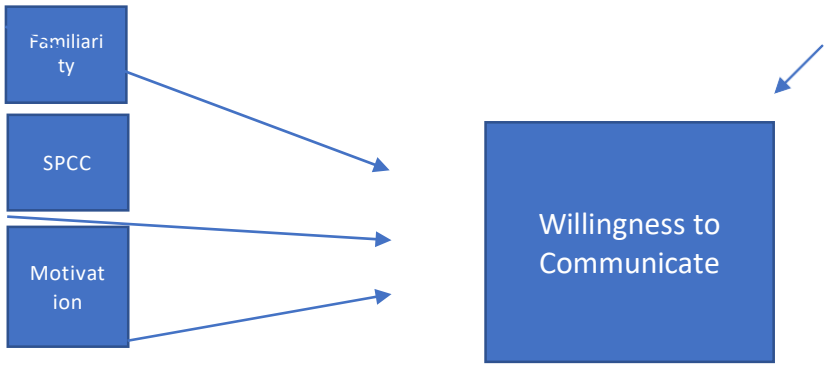
There are two main components of a network analysis. First is that it has edges which are basically the paths that were just mentioned. Secondly, it has junctions which represent points where edges meet. They can also mean valves in case a network has valves. In addition, both edges and junctions have the characteristics as enabled which allow flow and disabled which do not allow flow.

Network analysis has various applications such as transportation network, utility network and such where different lines and connections link data sets, vectors, systems networks and so much more.

1.4 Willingness to Communicate

In the research on Willingness to Communicate (WTC) in English as second language (L2) where a research protocol was designed for a meta-analysis procedure, prediction being its ultimate outcome requires the use of modeling techniques. As such, Structural Equation Model (SEM) constitutes a robust understanding and interpretation of predictor variables (familiarity, self-perceived communication competence and motivation) due to causality apart from correlation of variables. For instance, the presence of integrativeness can cause familiarity to rise giving way to WTC as a positive outcome. Integrativeness is a mediating variable, a latent variable that explains why familiarity is correlated to WTC. In another scenario, familiarity as observed variable can also explain why self-perceived communication competence (SPCC), another predictor variable as observable which means that familiarity can become a latent variable. The same analysis holds true for other predictor variables that are discussed more thoroughly in the next sections.

**Model 1. A Simple SEM of WTC**




According to Model 1, Willingness to Communicate (WTC) as the criterion or dependent variable is predicted to have some variation based on how each of the predictor variables – familiarity, SPCC, and motivation affect WTC indicated by arrow directions. What this means is that, familiarity, for instance, as an observed variable that can be numerically identified based on a number of communicators that affect WTC as to how familiar a person or a group of person will be, explains that a direct correlation between familiarity and WTC will be achieved computed by the correlation coefficient. The same situation is true for SPCC and motivation that explain correlative relationship with WTC. The arrow on the top right of WTC is a residual variable otherwise called an 'error term' that explains that variability of WTC can be explained aside from the given predictor variable. If treated as a single predictor variable, the linear regression equation will be as follows:

$$Y = bo + b1x1 + E$$

Where Y is WTC, the criterion variable determined by bo as the Y-intercept which is the height of the line fit that determines the linear regression and b1 is the slope of the line and x1 is familiarity as predictor variable. For instance, if the slope of the line is .80, what it indicates is that 80% of the variation in WTC is explained by the variation in familiarity.

Since there are two other predictor variables, SPCC and motivation, the same pathway analysis can be established such that SPCC will have a direct correlation to WTC and the following multiple linear regression equation is developed:

$$Y = bo + b1x1 + b2x2 + E$$

where (bo + b1x1) is the same simple linear equation discussed previously and b2 is the slope of SPCC while x2 is SPCC, another predictor variable and E as the error term. In this equation, two independent variables are now explaining the variation in WTC signifying a direct relationship for each predictor variable. The inclusion of motivation as another predictor variable completes the multiple linear regression equation:

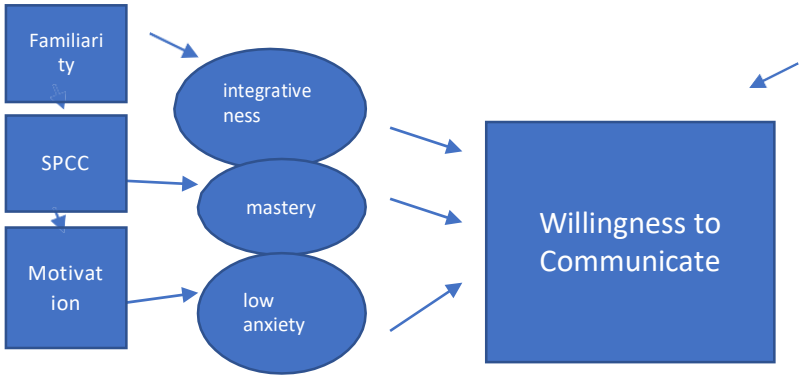
$$Y = bo + b1x1 + b2x2 + b3x3 + E$$

where (bo + b1x1 + b2x2 +) has been discussed previously and b3 represents the slope of motivation line as it affects variation in WTC and x3 is motivation. What the final multiple linear regression means is that each predictor variable (familiarity, SPCC and motivation) will predict the level of WTC.

However, meta-analysis protocol in previous submission indicates the presence of relationship between predictor variables instead of a direct relationship to the criterion variable. Having a said, another model can be developed following the presence of such structured relationship among predictor variables.

Finally, it can be noticed that the path diagram in Model 1 may be considered a path analysis based on correlative relationship established between variables with the absence of latent variables. However, the presence of multiple linear regression equations creates the essence of following STEM analysis rather than a simply path diagram

**Model 2. SEM Analysis with Latent Variables**




Model 2 highlights structural equation model (SEM) where structures are imposed on relationships among predictor variables instead of a unique relationship between predictor and criterion variable only. For instance, familiarity may affect self-perceived communication (SPCC) which in turn affects the level of WTC. What this indicates is a structured model where prediction of WTC is likely to be the result of two mediating variables – familiarity and SPCC. Next, SPCC might in turn affect the level of motivation that leads to a person's WTC as seen in the study among Japanese students in the previous meta-analysis protocol. The combination of Model 1 and Model 2 determines progress of correlation more than just each predictor variable unto its criterion variable but the combined effects of predictor variables can have more meaningful explanation to the variation of effects.

Next, encircled mediating variables (integrativeness, mastery and low anxiety) provide causality for respective predictor variables. For instance, familiarity is due to integrativeness of a person where ability to communicate becomes hassle-free and comfortable, thus, WTC will have positive outcome. Next, language mastery is perceived to cause self-perceived communication competition (SPCC) with the idea that proficiency in English as a second language increases desire to communicate such language, thus, WTC will have positive predictive outcome. Lastly, low levels of anxiety meaning less fear can influence higher motivation to communicate, thus, WTC leads to high prediction.

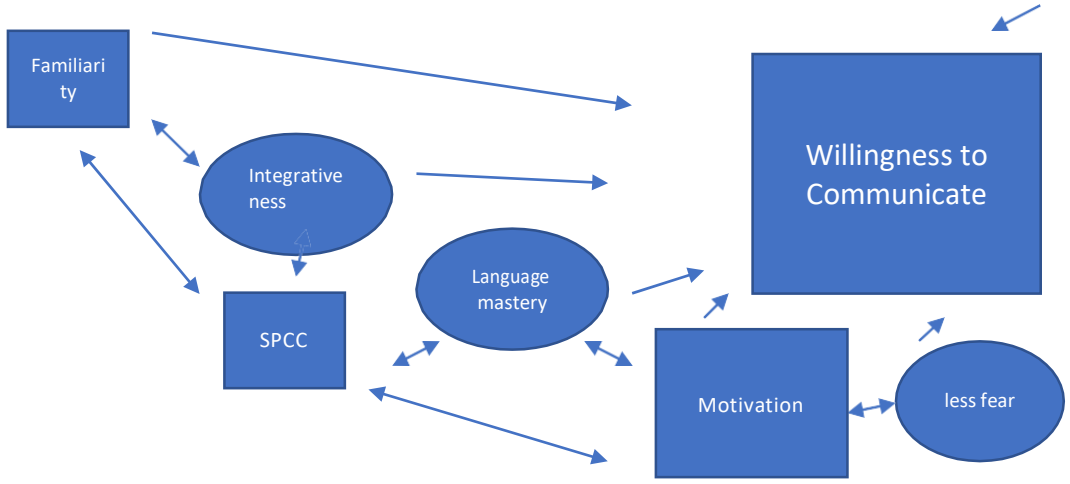
In another view, familiarity as observed variable can also be a mediating variable for SPCC which means familiarity causes SPCC to rise, vice versa. This is seen in the two-way arrow indication between said variables. The same explanation applies to motivation as on observed variable can also explain why SPCC happens, thus, it becomes a latent mediating variable in this case.

The above SEM provides robust understanding of the nature of the predictor variables including their causes of existence that leads to variability in WTC as outcome. The presence of mediating variables attests to the causality that is established which a simple multiple linear regression fails to establish as its focus is mainly correlational relationship.

1.4 Mediating Variables

The importance of causality highlights the presence of mediating variables. In this instance, improvement upon correlational analysis to determine causality explains high credibility of pathway analysis by structuring causality: does familiarity cause WTC to increase? Does SPCC increase WTC? Does motivation cause WTC? Other forms of questioning can be asserted using the following upgraded model with mediating variables using the same predictor variables but arranged differently to include mediating structures:

**Model 3. Upgraded SEM**




What Model 3 shows is a more complex structure of relationships more than a simple correlation between predictor and criterion variables only. The different arrows indicate many forms of mediating variables underplay that have multiple interpretations of given data. For instance, familiarity may influence SPCC and affect WTC while SPCC can also affect familiarity that will influence WTC and motivation affects SPCC and familiarity which influences WTC. In another instance, familiarity will influence SPCC which influences motivation and the combination of three instances will predict values for WTC. The arrow on the top right of WTC in the model indicates the presence of a residual variable which can also be called 'error term' indicating another variability of WTC that is not determined in the predictor variables. Model 3 offers many scenarios of mediating variables which can be the predictor variables themselves arranged differently according to the desired analysis.

Note the presence of mediating variables: integrativeness perceived to cause familiarity; mastery of language influencing SPCC and less fear contributing to motivation. All three mediating variables explain why predictor variables exist. In other words, these are the latent variables (in circles) that explain unobserved variables not measured as observed variables would normally do. Also in the model observed variables are enclosed in squares indicating measurable observed variables as the definition implies.

In another view, integrativeness as a mediating variable causing familiarity can also influence SPCC as indicated by a two-way arrow; less fear as contributory to motivation can also lead to SPCC and so on. With the presence of interlocking, multi scenarios of relationships, SEM provides deeper understanding more than path analysis or correlational regression analysis.

1.5 Moderating Variable

Based on original objective of determining how each of the predictor variables (familiarity, SPCC and motivation) will affect WTC, the moderating variable in the study will be population size (number of participants). This is due to the idea that higher number of participants will increase the effect of familiarity, SPCC and motivation on WTC in given contexts, the latter will increase its value. This is due to the correlation coefficient that will likely increase as a result of increasing data on participants. However, the mere presence of moderating variable does not explain the reason or reasons why such correlational relationship exists between predictor and criterion variable. For instance, multiple regression result can predict values of WTC given Pearson's square coefficient but it does not establish causality between variables.

1.6 Multiple Regression

The purpose of running multiple regression is to establish predictability of WTC based on how each of the predictor variables familiarity, SPCC and motivation affect WTC which answers the research question as specified in earlier submission of WTC research protocol: What is the degree of correlation of familiarity, SPCC and motivation on the level of WTC in English as L2 speakers using Pearson's correlational coefficient? From this result, research question 2 will be answered: Which variable has the highest degree of correlation to WTC.

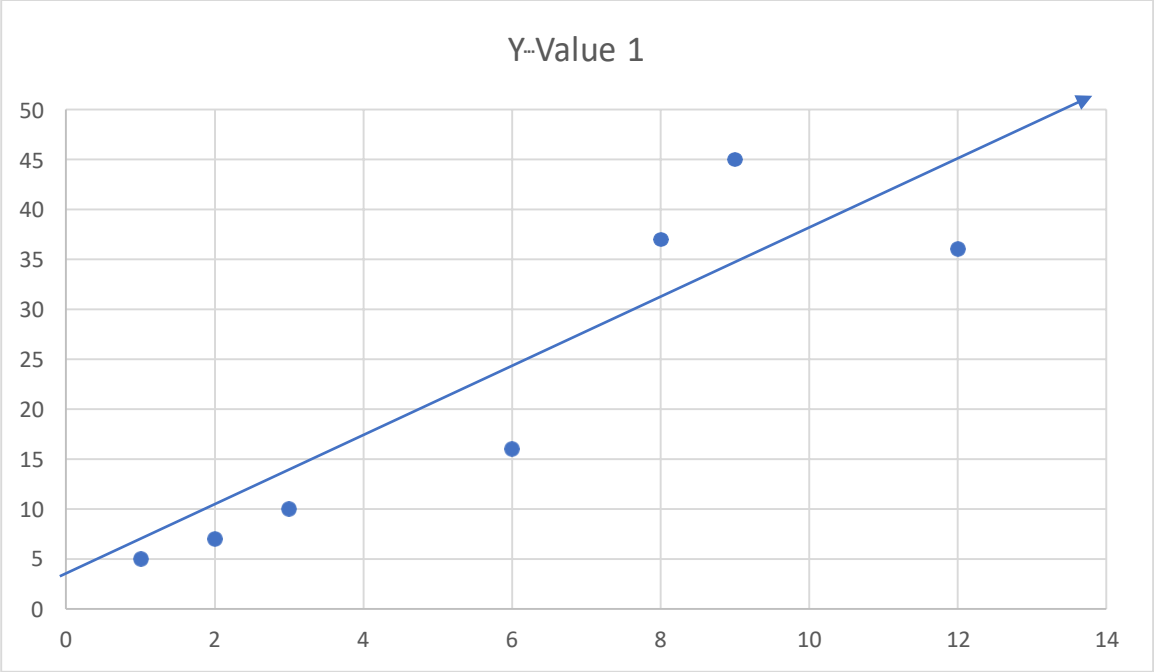
Clearly, multiple regression is required to determine predictability of WTC. This was earlier established in the following equation:

$$Y = bo + b1x1 + b2x2 + b3x3 + E$$

where:

Y = WTC
bo = Y-intercept determines the height of the regression line as it intersects the Y axis
x1 = familiarity
x2 = SPCC
x3 = motivation
b1 = slope of x1
b2 = slope of x2
b3 = slope of x3
E = error term

Based on earlier discussions, positive correlation exists between each predictor variable to WTC so an upward sloping regression line can be drawn to determine the positive relationship and scatter plots can also be done to determine degree of correlation.




The above hypothetical scatter plot indicates WTC on the Y-axis and familiarity on the x-axis. The dots indicate points of correlation between familiarity and WTC as indicated by correlation coefficient. Including SPCC and motivation will have multiple scatter plots of various predictor variables as each one affects WTC.

The upward sloping line is the regression line of the hypothetical case where the Y-intercept indicates a positive value higher than zero. Basically, it tells the height of the regression line and the scatter points are everywhere above or below the line. The closer the plots to the regression line the better the fit. In order to minimize the distance from points to the regression line, Ordinary Least Square Method will be done. The rationale behind this procedure is to determine the goodness of fit of the predictor variables of familiarity, SPCC and motivation as they affect WTC.

After going through OLS method where predicted values of correlation coefficient are determined thru multiple linear regression analysis, simple descriptive statistics such as mean for familiarity, SPCC and motivation will be determined as well as standard deviation for each including the combined mean for all three predictor variables as discussed in previous research protocol. The importance of reiterating descriptive statistics is its probability of increasing predictability apart from multiple regression given values for mean and standard deviation.

**Conclusion**

What the previous modeling techniques signify are various scenarios where predictability of WTC as criterion variable will be ascertained using pathway analysis in increasing levels – simple pathway analysis to determine prediction of WTC based on each predictor such as familiarity, SPCC and motivation and to a more structured relationship established among interaction of predictor variables in different arrow indications as they predict Willingness to Communicate (WTC) where this research can follow as deemed important to the researcher. Likewise, simple linear regression to multiple linear regression analyses indicate how each predictor variable will predict variations in WTC.

This paper concludes with the idea that modeling techniques offer different scenarios of prediction with mediating variables explaining causality of each of the predictor variable on WTC and between predictors themselves with combined effects on WTC, among others. Specifically, SEM is chosen to fit discussions of predictability that includes causality among predictor variables and mediating variables to WTC as criterion variable.

## References

Denis, D. (2021). Path analysis and structural equation modelling. https://onlinelibrary.wiley.com/doi/10.1002/9781119583004.ch15

Eric, M. Introduction to Networks. https://campus.datacamp.com/courses/introduction-to-network-analysis-in-python/introduction-to-networks?ex=1

Hanneman, R.A., & Riddle, M. (2011). Concepts and measures for basic network analysis. https://www.researchgate.net/publication/285702478_Concepts_and_measures_for_basic_network_analysis

Mike, A., Titsworth, S. & Hunt, S (2009). Advanced Modeling Techniques. SAGE Publications, Inc. https://books.google.com.ph/books/about/Quantitative_Research_in_Communication.html?id=xvt0AwAAQBAJ&redir_esc=y

Robinson, H. (2017). Network Analysis: Basic Concepts. Retrieved from www.geominds.com

Stein, C., Morris, N., & Nock, N. (2012). Structural equation modeling. https://www.researchgate.net/publication/221808236_Structural_Equation_Modeling

Stoelting, R. (2002). Structural Equation Modeling/Path Analysis. Retrieved from userwww.sfsu.edu.

…… Introduction to Multiple Regression. Retrieved from www.biddie.com

…… Path Analysis (statistics). Retrieved from www.wikipedia.org.