MJAL5:3 Spring2013 ISSN 0974-8741

A Review of Structural Equation Modeling (SEM) and its Application in Language Education Research

Akram Nayernia

Ph.D of Applied Linguistics, University of Tehran

nayernia@ut.ac.ir

Abstract

The present article reviews SEM as a statistical tool and its current use in language research, and especially in language assessment. First an introduction is given about the general features of this statistical tool. Its history and distinguishing characteristics are explored in the next sections. The steps of applying SEM are explained. Then, the application of SEM in language assessment is presented in time order.

Keywords: Structural Equation Modeling (SEM), Language learning, Language assessment

MJAL5:3 Spring2013 ISSN 0974-8741

A Review of Structural Equation Modeling (SEM) and its Application in Language Education Research by Akram Nayernia

Introduction

Structural equation modeling (SEM) is a sophisticated statistical methodology to testing hypotheses about relations among observed and latent variables (Hoyle, 1995:1) by modeling complicated functional or causal relationships among these variables. It is a powerful technique that can combine complex path or simultaneous equation models with latent variables measured by a factors analysis. In fact, it combines measurement model or confirmatory factor analysis (CFA) and structural model into a simultaneous statistical test (Hoe, 2008). Wright (1921) and Simon (1953) defined SEM as "a statistical technique for testing and estimating causal relations using a combination of statistical data and quantitative causal assumptions". The term *structural equation modeling* conveys two important aspects of the procedure: (a) that the causal processes under study are represented by a series of structural (i.e., regression) equations, and (b) that these structural relations can be modeled pictorially to enable a clearer conceptualization of the theory under study (Byrne, 2010: 3).

SEM is a general term that describes a large number of statistical models used to evaluate the consistency of substantive theories with empirical data. It represents an extension of general linear modeling procedures such as analysis of variance and multiple regressions. In addition, SEM can be used to study the relationships among latent constructs that are indicated by multiple measures and is applicable to experimental or non-experimental data and to cross-sectional or longitudinal data (Lei, 2007).

According to Gefen et al. (2000), SEM models consist of observed variables (also called manifest or measured) and unobserved variables (also called underlying or latent) that can be independent (exogenous) or dependent (endogenous) in nature. Latent variables are hypothetical constructs that cannot be directly measured, and in SEM are typically represented by multiple manifest variables that serve as indicators of the underlying constructs. The SEM model is an a priori hypothesis about a pattern of linear relationships among a set of observed and unobserved variables.

The use of SEM to examine complex questions in education and the social sciences has seen substantial growth in popularity over the past decade. The increase in use can be attributed to a number of issues including a greater flexibility in representing relationships among theoretical constructs, ability to posit latent constructs presumed to be underlying causes of observed manifest variables, and ease in evaluating the general compatibility or "goodness of fit" of a proposed model for the data being examined and the strength of relationships among constructs (Quintana & Maxwell, 1999).

As Kunnan (1998) puts it, SEM can be regarded as an integration of several models: multiple regression, path analysis and factor analysis. In the regression model, a directional relationship between two sets of measured variables, the dependent variable and a set of regressor variables, is hypothesized and evaluated; in the path analysis model theoretical relationships among independent measured variables and dependent measured variables are tested and the direct and indirect effects of the independent variables on the dependent variables are measured; the factor analysis model attempts to determine which observed measured variables share common variance– covariance characteristics with a latent construct or factor. SEM is a combination of these models which offers the mechanism to hypothesize relationships between constructs and measured variables and among constructs based on substantive theory (Lee, 2007).

In general, SEM involves two primary components: the measurement model and the structural model. The measurement model describes the relationships between observed variables and the construct or constructs to be measured. This model allows the researcher to determine how well the observed (measured) variables combine to identify underlying hypothesized constructs. In other words, measurement model deals with the relationship between observed and latent variables. Confirmatory factor analysis is used to test measurement model. In contrast, the structural model describes the

MJAL5:3 Spring2013 ISSN 0974-8741

interrelationships among constructs or latent variables, only. The interrelationships among latent variables are described as covariances, direct effects, or indirect (mediated) effects. The structural model is tested through regression models. When the measurement model and the structural model are considered together, the model may be called the composite or full structural model (Kunnan, 1998; Dilalla, 2000; Weston and Gore, 2007).

The History of SEM

Since SEM is a collection of related statistical techniques, it does not have a single source. Golob (2001) argues that SEM is the union of latent variable (factor analytic) approaches, developed primarily in psychology and sociology, and simultaneous equation methods of econometrics. In his view, modern SEM evolved out of the combined efforts of many scholars pursuing several analytical lines of research. To discuss the history of SEM, one must consider development of four types of related models, namely, regression, confirmatory factor, path, and structural equation models.

The first model involves linear regression models which emerged by creation of a formula by Karl Pearson in 1896 for the correlation coefficient. This formula provides an index for the relationship between two variables (Pearson, 1938). Regression models employ a correlation coefficient and least square criterion to compute regression weights. This model makes prediction of dependent observed variable scores possible given a linear weighting of a set of independent observed scores that minimizes the sum of squared residual values. Regression analysis provides a test of a theoretical model that may be useful for prediction (Schumacker & Lomax, 2004).

Some years later, Charles Spearman (1904, 1927 cited in Schumacker & Lomax, 2004) used the correlation coefficient to determine which items correlated or went together to create the factor model. His basic idea was that if a set of items correlated or went together, individual responses to the set of items could be summed to yield a score that

MJAL5:3 Spring2013 ISSN 0974-8741

would measure, define, or imply a construct. Spearman used the term factor analysis in defining a two-factor construct for a theory of intelligence for the first time. Some years later, in 1940, Lawley and Thurstone further developed applications of factor models and proposed instruments to yield observed scores from which constructs could be inferred. The term *confirmatory factor analysis* (CFA) as used today is based in part on works of Howe (1955), Anderson and Rubin (1956), and Lawley (1958). The CFA method was more fully developed by Karl Joreskog in the 1960s to test whether a set of items defined a construct (Schumacker & Lomax, 2004).

The third model to be considered in the history of SEM is path analysis developed by Sewall Wright, a geneticist, developed the basics of path analysis (1921, 1934), which was subsequently introduced by various authors to researchers in other disciplines such as sociology, economics, and later psychology. An annotated bibliography by Wolfle (2003) traces the introduction of path analysis to the social sciences. Path models use correlation coefficients and regression analysis to model more complex relationships among observed variables. In many respects, path analysis involves solving a set of simultaneous regression equations that theoretically establish the relationship among the observed variables in the path model (Schumacker & Lomax, 2004). Path analysis introduced three concepts: 1) the first covariance structure equations, 2) the path diagram or causal graph, and 3) decomposition of total effects between any two variables into total, direct, and indirect effects (Golob, 2000)

In the early 1970s, these approaches, that is, measurement (factor analysis) and structural (path analysis), were integrated in the work of basically three authors, K. G. Jöreskog, J. W. Keesling, and D. E. Wiley, into a framework called JKW model by Bentler (1980). SEM was initially popularized by the wide distribution of linear structural relationships (LISREL) program developed by Jöreskog (1970), Jöreskog et al. (1970), and Jöreskog and Sörbom (1979) as one of the first widely available computer programsto analyze models based on the JKW framework. More complete versions of LISREL became available in subsequent years. The 1980s and 1990s witnessed the development of many more computer programs and a rapid expansion of the use of SEM techniques in areas such as developmental psychology, behavioral genetics, sports medicine, education, and public health, etc (Kline, 2005).

Characteristics of SEM

In this section some defining features of SEM which set it apart from any other statistical technique are mentioned.

1. Computer programs used for SEM require a lot of information about variables assumed to be affecting other variables and the direction of these effects. These a priori specifications reflect the researchers' hypothesis, and constitute the model to be evaluated in the analysis. In this sense, SEM takes a confirmatory rather than an exploratory approach to data analysis. That is, one of the main questions to be answered by SEM analysis is whether the hypothesized model is supported by the data. However, as it is often the case, when the data is inconsistent with the model, i.e. does not support the model, the researcher must either abandon the model or modify the hypotheses on which it is based. If the hypotheses are modified to fit the data, the analysis gets a more exploratory nature, since the revised model is tested with the same data. Furthermore, by demanding that the pattern of relations among variables be specified a priori, SEM lends itself well to the analysis of data for inferential purposes where the pattern of inter-relationships among the study constructs are specified a priori and grounded in established theory. By contrast, most other multivariate procedures are essentially descriptive in nature (e.g., exploratory factor analysis), so that hypothesis testing is difficult, if not impossible (Kline, 2005; Hoe, 2008; Byrne, 2010).

Jöreskog (1993) expressed these ideas more formally by distinguishing among (1) strictly confirmatory, (2) alternative models, and (3) model-generating applications of SEM. The first refers to when the researcher has a single model and tests it with empirical data, which is accepted or rejected based on its correspondence with the data. The second context which may be more frequent than the first is restricted to situations

MJAL5:3 Spring2013 ISSN 0974-8741

where more than one a priori model is available. In these situations, the researcher specifies several alternative models and tests them with empirical data. The last context, model generating, is probably the most common and occurs when an initial tentative model does not fit the data and is modified on the basis of suggestions from SEM and substantive theory. The altered model is then tested again with the same data until a satisfactory model emerges. The goal of this process is more to "discover" a model with two properties: It makes theoretical sense, and its statistical correspondence to the data is reasonable.

2. While some standard statistical procedures (e.g. ANOVA, multiple regression) are based on observed measurements only and do not differentiate between observed and latent variables, methods using SEM procedures can incorporate both unobserved and observed variables (Kline, 2005; Byrne, 2010). In fact, SEM models are usually conceived as not directly measurable, and possibly not well-defined, theoretical or hypothetical constructs, such as anxiety, attitudes, intelligence, learning strategies, etc. (Raykov and Marcoulides, 2006).

Kline (2005) points to the characteristics of SEM in terms of the distinctions between latent and observed variables as following:

a. It is not necessary for models to have hypothesized latent variables. The evaluation of models that concern effects only among observed variables is also possible in SEM.
b. There is more than one type of latent variable, each of which reflects different assumptions about the relation between observed and unobserved variables.
c. Latent variables in SEM can represent a wide range of phenomena. For example, theoretical constructs about characteristics of persons (e.g., phonological processing, verbal reasoning), of higher-level units of analysis (e.g., characteristics of geographic regions), or of measures, such as method effects (e.g., parent vs. teacher informants), can all be represented as latent variables in SEM.

MJAL5:3 Spring2013 ISSN 0974-8741

d. The observed-latent distinction also provides a way to take account of imperfect score reliability. This is not to say that SEM can be used as a way to compensate for gross psychometric flaws. It, like any other technique, cannot compensate gross psychometric flaws, but this aspect of SEM can lend a more realistic quality to an analysis.

3. The general SEM system is estimated using covariance (structure) analysis, whereby model parameters are determined such that the variances and covariances of the variables implied by model system are as close as possible to the observed variances and covariances of the sample. In other words, the estimated parameters are those that make the variance–covariance matrix predicted by the model as similar as possible to the observed variance–covariance matrix, while respecting the constraints of the model (Golob, 2001). Covariance is the basic statistic of SEM and there are two main goals of the analysis: to understand patterns of correlations among a set of variables, and to explain as much of their variance as possible with the model specified by the researcher (Kline, 2005: 13). In other words, SEM models are usually fir to matrices of interrelationship indices, i.e. covariance or correlation matrices, between all pairs of observed variables and variables means (Raykov and Marcoulides, 2006).

4. SEM technique can be applied to both experimental and non-experimental data. Although there is common view that SEM is only appropriate for non-experimental data, SEM techniques can also be used in studies that have a matrix of experimental and non-experimental features (Kline, 2005).

5. The SEM family includes many standard statistical procedures, including multiple regression, canonical correlation, and the analysis of variance (ANOVA). All of these techniques are special instances of SEM referring to the broad generality of SEM. In fact, as Kline (2005) puts it, many statistical terminologies are merely conveniences that allow us to quickly associate something with SEM analysis.

MJAL5:3 Spring2013 ISSN 0974-8741

6. While traditional multivariate procedures are incapable of assessing or correcting for measurement error, SEM provides explicit estimates of these error variance parameters in all observed variables, particularly in the independent (predictor or explanatory) variables. In fact, alternative methods (e.g. those rooted in regression, or the general linear model) ignore errors in the independent variables, leading to inaccurate results, especially when the errors are significant. Whereas, such inaccuracies are avoided when corresponding SEM analyses are used. This is achieved by including an error term for each fallible measure, whether it is an explanatory or predicted variable. The variances of the error terms are, in general, parameters that are estimated when a model is fit to data. Tests of hypotheses about them can also be carried out when they represent substantively meaningful assertions about error variables or their relationships to other parameters (Byrne, 2010; Raykov and Marcoulides, 2006).

Steps in SEM

Bollen and Long (1993) list the basic steps to be followed in the application of SEM as model specification, model identification, model estimation, testing model fit, and model modification or model manipulation. They are actually iterative because problems at a later step may require a return to an earlier one. These steps are briefly explained below.

1. Model specification. The first step in SEM analysis is the specification of a model to be estimated. In fact, at this stage, the researcher's hypotheses are formulated as a structural equation model. Model specification involves using all available relevant theory, research and information, and developing a theoretical model. In other words, available information is used to specify the variables to be included in the theoretical model and their interrelationships. As Schreiber (2008) points out the failure to include relevant variables will create a specification error which will in turn cause the estimations to be incorrect and therefore result in inappropriate inferences. As noted above, an SEM model generally consists of two parts, the *measurement model* and the *structural model*. The measurement model specifies the relationships between measured variables and latent variables that are not directly measurable but are specified, and the structural model specifies the direct and indirect relationships among the latent variables.

Model specification involves formulating a statement about a set of parameters. The parameters to be specified are constants that indicate the nature of the relation between two variables. Parameters are usually specified as fixed, free, or constrained. *Fixed parameters* are not estimated from the data and are typically fixed at zero or one (indicating no relationship between variables). The paths of fixed parameters are labeled numerically (unless assigned a value of zero, in which case no path is drawn) in a SEM diagram. *Free parameters* are unknown parameters estimated from the observed data and believed by the investigator to be non-zero. Constrained parameters are also unknown but constrained to equal one or more other parameters. Determining parameters in a SEM analysis is extremely important because it determines which parameters will be used to compare the hypothesized diagram with the sample population variance and covariance matrix in testing the fit of the model (Step 4). The choice of which parameters are free and which are fixed in a model is up to the researcher. This choice represents the researcher's a priori hypothesis about which pathways in a system are important in the generation of the observed system's relational structure (e.g., the observed sample variance and covariance matrix) (Hoyle, 1995; Schumacker and Lomax, 2004).

The relationships among variables, both observed and latent, can be described as association or covariance, direct effect, and indirect or mediated effect. Covariances are analogous to correlations in that they are defined as non-directional relationships among variables. They are pictorially depicted using doubleheaded arrows in path diagrams. Direct effects, "the building blocks of structural equation models" (Hoyle, 1995: 3) are directional relationships among measured and latent variables, similar to those found in ANOVA and multiple regressions. These relations are indicated by single-directional arrows. An indirect effect is the relationship between an independent variable and a dependent variable that is mediated by one or more variables (Baron and Kenny, 1986).

MJAL5:3 Spring2013 ISSN 0974-8741

Mediation may be full or partial. That is to say, indirect effects indicate the effect of one independent variable upon on dependent variable through one or more intervening or mediating variable(s).

2. Model identification. In broad terms, the issue of identification focuses on whether or not there is a unique set of parameters consistent with the data (Byrne, 2010: 33). According to Hoyle (1995) identification refers to the correspondence between the free parameters and the observed variances and covariances. It concerns whether a single, unique value for each and every parameter can be obtained from the observed data. As Schumacker and Lomax (2004) point out, 'model identification depends on the specification of parameters as free, fixed or constrained. Once the model is specified and the parameter specifications are indicated, the parameters are combined to form one and only one ∑ (model-implied variance–covariance matrix)' (p. 81). The question to be asked in this stage is: Is it possible to find a unique set of parameter estimates on the basis of the sample data contained in the sample covariance matrix S and the theoretical model implied by the population covariance matrix ∑ (Schumacker and Lomax, 2004)?

Structural models may be *just-identified*, *overidentified*, or *underidentified*. A just-identified model is one in which there is a one-to-one correspondence between the data and the structural parameters. In other words, the number of data variances and covariances equals the number of parameters to be estimated. In this situation, a value can be obtained through one and only one manipulation of the data for each parameter. However, despite the capability of the model to yield a unique solution for all parameters, the just-identified model is not scientifically interesting because it has no degrees of freedom and therefore can never be rejected. An overidentified model is one in which the number of parameters to be estimated is less than the number of data points (i.e., variances and covariances of the observed variables). That is, for one or more free parameters, a value can be obtained in multiple ways from the observed data. This situation results in positive degrees of freedom that allow for rejection of the model, thereby rendering it of scientific use. The aim in SEM, then, is to specify a model such

MJAL5:3 Spring2013 ISSN 0974-8741

that it meets the criterion of overidentification. Finally, an underidentified model is one in which the number of parameters to be estimated exceeds the number of variances and covariances (i.e., data points). In other words, it is not possible to obtain a single, unique value for one or more free parameters from the data. As such, the model contains insufficient information (from the input data) for the purpose of attaining a determinate solution of parameter estimation; that is, an infinite number of solutions are possible for an underidentified model (Hoyle, 1995; Byrne, 2010).

3. **Model estimation**. According to Iriondo et al. (2003) the aim of this stage is to estimate the value of the unknown parameters, such as the standardized path coefficients, in such a way that the observed variance–covariance matrix is optimally adjusted to the predicted moment matrix. "Estimation concerns the procedure to be used to derive the parameter estimates, such as the coefficients and standard errors" (Schreiber, 2008). Estimation involves determining the value of the unknown parameters and the error associated with the estimated value from a set of observed data. It is desired to obtain estimates for each of the parameters specified in the model that produce the implied matrix ∑, in a way to yield a matrix as close as possible to S, sample covariance matrix of the observed variables.

In this process, a particular *fitting function* is used to minimize the difference between ∑ and S. Severla fitting functions or estimation procedures are avaialbe. Some of the earlier methods include unweighted or ordinary least squares (ULS or OLS), generalized least squares (GLS), and maximum likelihood (ML). GLS and ML are preferred over single stage least square method such as those used in standard ANOVA or multiple regression. More recently, other estimation procedures have been developed for the analysis of covariance structure models. Automatic starting values - for LISREL - have been provided for all of the parameter estimates. These are referred to as *initial estimates* and involve a fast, noniterative procedure, unlike other method such as ML, which is iterative. Iterative methods involve a series of attempts to derive estimates of free parameters that imply a covariance matrix like the observed one (Hoyle, 1995; Schumacker and Lomax, 2004).

4. Testing model fit. Once the parameter estimates are obtained for a specified model, it must be determined how well the data fit the model. Evaluation of model fit concerns the extent to which the obtained sample data support the theoretical model. According to Hoyle (1995), a model fits the observed data to the extent that its implied covariance matrix is equivalent to the observed covariance matrix (that is, the elements of the residual matrix are near zero). The more free parameters in a model the more likely the model is to fit the data, because parameter estimates are derived from the data. Besides, the effectiveness of different estimation methods depends upon the sample size and model complexity.

As Schumacker and Lomax (2004) point out, model fit can be viewed from two perspectives. The first one is to consider some global test of fit for the whole model. The second one is to examine the fit of individual parameters of the model. The global tests in SEM are known as model fit criteria. Unlike many statistical procedures that have a single, most powerful fit index (e.g., *F* test in ANOVA), in SEM has a large number of model fit indices. Many of these measures are based on a comparison of the model-implied covariance matrix to the sample covariance matrix. If these two covariances are similar to some extent, it can be concluded that the data fit the theoretical model. If they are quite different, then it can be argued that the data do not fit the theoretical model. Regarding the second perspective, three main features of the individual parameters can be considered. One feature is whether a free parameter is significantly different from zero. Once parameter estimates are obtained, standard errors for each estimate are also computed. A ratio of the parameter estimate to the estimated standard error can be formed as a *critical value,* which is assumed normally distributed (unit normal distribution), that is, critical value equals parameter estimate divided by standard error of the parameter estimate. If the *critical value* exceeds the expected value at a specified *a* level, then that parameter is significantly different from zero. The

MJAL5:3 Spring2013 ISSN 0974-8741

parameter estimate, standard error, and critical value are routinely provided in the computer output for a model. A second feature is whether the sign of the parameter agrees with what is expected from the theoretical model. A third feature is that parameter estimates should make sense, that is, they should be within an expected range of values. In sum, all free parameters should be in the expected direction, be statistically different from zero, and make practical sense.

Multiple indices are available to evaluate model fit. The most stringent concept of fit suggests that the model must exactly replicate the observed data. A second perspective is that models approximating the observed data are acceptable. Hoyle and Panter (1995) have recommended that researchers report several indices of overall model fit. In the following fit indices reported by most software programs are introduced.

GFI and X^2. Absolute fit indices directly assess how well a model fits the observed data and are useful in comparing models when testing competing hypotheses. Absolute fit indices include the goodness-of-fit index (Jöreskog & Sörbom, 1981), X^2 (Bollen, 1989), and scaled X^2 (Satorra & Bentler, 1994). GFI is similar to R_2, which is used in regression to summarize the variance explained in a dependent variable, although GFI refers to the variance accounted for in the entire model. However, GFI is not reported as consistently as X^2. X^2 is directly derived from the value of fitting function, that is, it results from the value of fitting function and the sample size minus one, $F(N\text{-}1)$ (Hoyle, 1995). X^2 values are actually tests of model misspecification. Hence, a significant X^2 suggests that the model does not fit the sample data. In contrast, a nonsignificant X^2 is indicative of a well model fit with the data. Despite the common use of X^2 as the absolute fit index, it suffers from two limitations. First, this statistic tests whether the model is an exact fit to the data and an exact fit is rarely found. Second, as with most statistics, large sample sizes increase power, resulting in significance with small effect sizes (Henson, 2006). Consequently, a nonsignificant X^2 may be unlikely, although the model may be a close fit to the observed data. Despite these limitations, researchers

MJAL5:3 Spring2013 ISSN 0974-8741

report the X^2 almost universally (Martens, 2005), and it provides a means for testing whether two models differ in their fit to the data.

CFI. Bentler's (1990) CFI is an example of an incremental fit index. This type of index compares the improvement of the fit of the researcher's model over a more restricted model, called an independence or null model, specifying no relationships among variables. CFI ranges from 0 to 1.0, with values closer to 1.0 indicating better fit.

RMSEA. The RMSEA is also recommended as a fit index (Steiger, 1990; Steiger & Lind, 1980). This index corrects for a model's complexity. As a result, when two models explain the observed data equally well, the simpler model will have the more favorable RMSEA value. A RMSEA value of .00 indicates that the model exactly fits the data. A recent practice is to provide the 90% CI as well for the RMSEA, which incorporates the sampling error associated with the estimated RMSEA.

SRMR. The SRMR (Bentler, 1995) index is based on covariance residuals, with smaller values indicating better fit. The SRMR is a summary of how much difference exists between the observed data and the model. The SRMR is the absolute mean of all differences between the observed and the model-implied correlations. A mean of zero indicates no difference between the observed data and the correlations implied in the model; thus, an SRMR of 0.00 indicates perfect fit.

5. Model modification. Rarely is a proposed model the best-fitting model. Consequently, modification (respecification) may be needed. This involves adjusting the estimated model by freeing (estimating) or setting (not estimating) parameters. Modification is a controversial topic, which has been likened to the debate about post hoc comparisons in ANOVA (Hoyle, 1995). As Martens (2005) reports, researchers generally accomplish modification by using statistical search strategies (often called a specification search) to determine which adjustments result in a better-fitting model. The Lagrange Multiplier test identifies which of the parameters that the researcher

MJAL5:3 Spring2013 ISSN 0974-8741

assumed to be zero are significantly different from zero and should be estimated. The Wald test, in contrast, identifies which of the estimated parameters that were assumed to be significantly different from zero are not and should be removed from the model. Schumacker and Lomax (2004) and Kline (2005) provide detailed information on conducting specification searches using modification indices.

SEM in language assessment

Kunnan (1998) lists the objectives of using SEM in language assessment as following:

1. Research on the exploration of the two-part conceptualization of construct validation of test score-use in order to improve test design: construct representativeness (components, processes and knowledge structures that are involved in test responses; and nomothetic span (relationship of the test to other measures of individual differences);
2. Research on the exploration of the factor structure of test performance or questionnaires in order to better understand the abilities assessed by tests or test taker characteristics collected through questionnaires of homogeneous and heterogeneous groups of test takers or respondents (examples, Muthen, 1989; Kunnan, 1995; Purpura, 1996; Ginther and Stevens, 1998 cited in Kunnan, 1998);
3. Research on the exploration of the hypothesized relationships among test taker characteristics or background (or external factors), test taking strategies and test performance in a second or foreign language context to better understand the effect of salient test taker characteristics on test performance (examples, Muthen, 1988, 1989; Kunnan, 1995; Purpura, 1996 cited in Kunnan, 1998);
4. Research on the exploration of hypothesized relationships among test task characteristics and test performance in order to better understand the effect of different test tasks (multiple methods) on test performance. SEM would provide a more powerful mechanism for this type of investigation than regression, which

has been previously used by researchers (examples, Freedle and Kostin, 1993; Bachman, Davidson and Milanovic, 1996 cited in Kunnan, 1998); and

5. Research on the exploration of population heterogeneity among test takers, since this is generally typical of most data sets (including language assessment data sets, especially in large-scale high stakes ESL/EFL tests). For example, in instructional or language assessment settings, widely varying curricula, opportunities to learn, exposure or instruction in target language may require data to be analyzed as independent multi-samples (example, Kunnan, 1995 cited in Kunnan, 1998) as well as simultaneous multi-samples (example, Ginther and Stevens, 1998 cited in Kunnan, 1998).

Bachman and Palmer (1981, 1982, and 1989) were the first authors to employ SEM in their studies on respectively construct validation of the FSI Oral Interview, components of communicative proficiency, and self-ratings of communicative language ability.

Other researchers utilizing SEM in their studies included Swinton and Powers (1980), who examined the component abilities that underlie performance on the TOEFL, Purcell (1983), who investigated models of pronunciation accuracy, Fouly (1985), who investigated the relationships among learner variables and second language proficiency, Wang (1988), who investigated cognitive achievement and psychological orientation among language minority groups, Hale et al. (1989), who studied the factor structure of the TOEFL, and Turner (1989), who investigated second language cloze test performance.

During the 1980s, Gardner and other second language acquisition researchers were using SEM with second language acquisition data (Gardner, et al., 1983; Gardner *et al.*, 1987; Gardner, 1988; Clement and Kruidenier, 1985; Ely, 1986) to investigate motivation, aptitude, and attitude as factors that affect second language acquisition. In 1984, Nelson et al. constructed and empirically evaluated a model of second language acquisition for adult learners. Their proposed structural equation model described the

MJAL5:3 Spring2013 ISSN 0974-8741

relationships between latent variables representing sociocultural background, cognitive ability (in the first language), functional language proficiency, cognitive language proficiency, attitudes, motivation, and instructional approach. Their results showed that an "integrative" approach to second language instruction was more effective than a strictly "behaviorist" approach, and functional language ability was an important component of the language acquisition process.

SEM applications in 1990s include Sasaki (1993), who investigated the relationships among second language proficiency, foreign language aptitude, and intelligence, Kunnan (1995, cited in Kunnan, 1998), who investigated the influence of some test taker characteristics on test performance in tests of English as a foreign language, Purpura (1996), who investigated the relationships between test takers' cognitive and metacognitive strategy use and second language test performance, and Ginther and Stevens (1998), who investigated the factor structure of an advanced placement Spanish language examination among four different Spanish-speaking test taking groups. Gardner et al. (1999) made use of SEM to direct attention to the role of early environmental characteristics and language learning motivation on subsequent language attitudes and perception of L2 competence.

Recently, SEM has been extensively used in studies on language learning and language assessment. Geldren et al. (2003) analyzed the relationship between L1, L2, and L3 reading comprehension and its constituent skills using SEM approach. Schoonen (2005) investigated the effect of writing proficiency, topic of writing assignment, the features of writing to be scored and scoring methods on writing score through structural equation modeling in a generalizability study using variance analytic techniques. Shin (2005) investigated the relationship between examinee proficiency and the structure of the Test of English as a Foreign Language (TOEFL) and the Speaking Proficiency in English Assessment Kit (SPEAK) using multi-group structural equation modeling.

Phakiti (2006) conducted an empirical study investigating the nature of cognitive and metacognitive strategies and their direct and indirect effects on EFL reading tests

MJAL5:3 Spring2013 ISSN 0974-8741

performance employing SEM. In'nami (2006) investigated the effects of test anxiety on listening test performance.

Shiotsu and Weir (2007), in a componential approach to modelling reading ability, investigated the contribution of knowledge of syntax and knowledge of vocabulary to L2 reading using SEM.

Phakiti (2008) tested a fourth-order factor model of strategic competence through the use of structural equation modeling (SEM). He examined the hierarchical relationship of strategic competence to (a) strategic knowledge of cognitive and metacognitive strategy use in general and (b) strategic regulation of cognitive and metacognitive strategy use in a specific second language reading test over a period of 2 months. Song (2008) employed SEM analysis to investigate if there are divisible sub-skills in L2 reading and listening comprehension. Hiromori (2009) examined a process model of L2 learners' motivation through SEM.

Papi (2010) examined a theoretical model that subsumed the ideal L2 self, the ought-to L2 self, and the L2 learning experience, English anxiety, and intended effort to learn English in an Iranian context through SEM. Lesaux et al. (2010) examined English reading comprehension skill development. They, specifically, investigated the effects of Spanish (L1) and English (L2) oral language and word reading skills on reading comprehension through SEM. Schroeders et al. (2010), utilizing SEM, examined the constructs of reading, listening, and viewing comprehension in an EFL context. Bae and

Rivers (2011) examined the four psychological facets of Japanese national identification in relation to a selection of English language learning processes by creating a structural model of causality. Pae and Shin (2011) examined the effects of differential instructional methods on the relationships among intrinsic and extrinsic motivations, self-confidence, and English as a foreign language (EFL) achievement using SEM.

MJAL5:3 Spring2013 ISSN 0974-8741

Conclusion

As a theory-testing tool, SEM is a promising research tool for researchers in the field of language education and language assessment. SEM's data processing ability may, for example, help researchers explore theories of language learning and acquisition, test theoretical models on language learning, explore the relationships among diverse factors contributing to language learning, explore the nature of language proficiency, performance on language tests, etc.

This article reviewed SEM as a statistical tool and its current use in language research, and especially in language assessment. Future studies should address in-depth how effectively SEM has been used in each of the cases, its appropriateness to the questions posed by researchers, and how SEM aids in testing the research hypotheses.

References

Bachman, L.F. & Palmer, A. (1981). The construct validation of the FSI oral interview. *Language Learning* 31, 67–86.

Bachman, L.F. & Palmer, A. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly* 16, 449–65.

Bachman, L.F. & Palmer, A. (1989). The construct validation of self-ratings of communicative language ability. *Language Testing* 6, 14–29.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

MJAL5:3 Spring2013 ISSN 0974-8741

Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin, 107,* 238-246.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research, 17,* 303-316.

Bollen, K.A. and Long, J.S. (1993). *Testing structural equation models*. Newbury Park, CA: Sage.

Byrne, B. (2010). *Structural equation modeling with AMOS: basic concepts, applications, and programming (2nd ed.)*. Routledge Taylor & Francis Group

Clement, R. and Kruidenier, B.G. (1985). Aptitude, attitude and motivation in second language proficiency: a test of clement's model. *Journal of Language and Social Psychology* 4, 21–37.

Dilalla, L. F. (2000) Structural equation modeling: uses and issues. In handbook of applied multivariate statistics and mathematical modeling. American press 439 -464

Ely, C.M. (1986). Language learning data: a description and causal analysis.
Modern Language Journal 70, 28–35.

Fouly, K. (1985). A confirmatory multivariate study of the nature of second language proficiency and its relationships to learner variables. Unpublished Ph.D. dissertation, University of Illinois, Urbana.

Gardner, R.C. (1988) The socio-educational model of second language learning: assumptions, findings and issues. *Language Learning* 38, 101–26.

Gardner, R.C., Lalonde, R.N., Moorcraft, R. & Evers, F.T. (1987) Second language attrition: the role of motivation and use. *Journal of Language and Social Psychology* 6, 1–47.

Gardner, R.C., Lalonde, R.N. & Pierson, R. (1983) The socio-educational model of SLA: an investigation using LISREL causal modeling. *Journal of Language and Social Psychology* 2, 1–15.

Gardner, R. C., Masgoret, A. & Tremblay, P. F. (1999). Home background characteristics and second language learning. *Journal of Language and Social Psychology*, 18, 4. 419 – 437

Gefen, D., Straub, D.W., Boudreau, M., 2000. Structural equation modeling and regression: guidelines for research practice. Communications of the AIS 1 (7), 1–78.

Geldren, A. V., Schoonen, R., Glopper, K., Hulstijn, J., Snellings, P., Simis, A., & Stevenson, M. (2003). Roles of linguistic knowledge, metacognitive knowledge and processing speed in L3, L2 and L1 reading comprehension: A structural equation modeling approach. *International Journal of Bilingualism*, 7, 1., 7 – 25.

Ginther, A. and Stevens, J. (1998). Language background, ethnicity, and the internal construct validity of the Advanced Placement Spanish language examination. In Kunnan, A.J., editor, *Validation in language assessment*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 169–94.

Golob, T. F. (2001). Structural equation modeling for travel behavior research. *Transportation Research Part B.* 37. 1-25

Hale, G.A., Rock, D.A. & Jirele, T. (1989) *Confirmatory factor analysis of the TOEFL*. TOEFL Research Report 32. Princeton: Educational Testing Service.

MJAL5:3 Spring2013 ISSN 0974-8741

Henson, R. K. (2006). Effect-size measures and meta-analytic thinking in counseling psychology research. *The Counseling Psychologist*, *34*, 601-629.

Hiromori, T. (2009). A process model of L2 learners' motivation: from the perspectives of general tendency and individual differences. *System*, 37. 313 – 321.

Hoe, S. L. (2008). Issues and procedures in adopting structural equation modeling technique. *Journal of Applied Quantitative Methods. Vol. 3, no. 1. 76 – 83*

Hoyle, R. H. (1995). *Structural Equation Modeling, Concepts, Issues, and Application.* Sage Publication.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling* (pp. 158-176). Thousand Oaks, CA: Sage.

In'nami, Y. (2006). The effects of test anxiety on listening test performance. *System*. 34. 317 – 340.

Iriondo, J. M., Albert, M. J., & Escudero, A. (2003). Structural equation modeling: an alternative for assessing causal relationships in threatened plant populations. *Biological Conservation*, 113, 367 – 377.

Jöreskog, K.G., Sörbom, D. (1996). LISREL 8: User’s Reference Guide. *Scientific Software International Inc*., Chicago, IL.

Jöreskog, K.G. (1970). A general method for analysis of covariance structures. *Biometrika* 57, 239–251.

Jöreskog, K.G., Gruvaeus, G.T., van Thillo, M. (1970). *ACOVS: A General Computer Program for Analysis of Covariance Structures*. Educational Testing Services, Princeton, NJ.

MJAL5:3 Spring2013 ISSN 0974-8741

Jöreskog, K.G., Sörbom, D. (1979). Advances in Factor Analysis and Structural Equation Models. Abt Books, Cambridge, MA.

Jöreskog, K. G., & Sörbom, D. (1981). *Analysis of linear structural relationships by maximum likelihood and least squares methods* (Research Report No. 81-8). Uppsala, Sweden: University of Sweden.

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Lang (Eds.), *Testing structural equation models* (pp. 294–316). Newbury Park, CA: Sage.

Kline, R. B. (2005). Principles and practice of structural equation modeling. The Guilford Press.

Kunnan, A. J. (1998). An introduction to structural equation modeling for language assessment research. *Language Testing 15*, 3: 295-332.

Lee, S. Y. (2007). *Structural equation modeling, a bayesian approach*. John Wiley & Sons, Ltd.

Lei, P. (2007). Structural Equation Modeling. In Salkind, N. J. and Rasmussen, K (Eds.) *Encyclopedia of Measurement and Statistics.* SAGE Publications.

Leohlin, J. C. (2004). *Latent Variable Models, An Introduction to Factor, Path, and Structural Equation Analysis* (4th ed.). Lawrence Erlbaum Associates, Publishers.

Martens, M. P. (2005). The use of structural equation modeling in counseling psychology research. *The Counseling Psychologist, 33,* 269-298.

Muthen, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics 22*: 43-65.

Nelson, F. H., Lomax, R. G. & Perlman, R. (1984). A structural equation model of second language acquisition for adult learners. *The Journal of Experimental Education*. 53, 1. 29 – 39.

Pae, T. & Shin, S. (2011). Examining the effects of differential instructional methods on the model of foreign language achievement. *Learning and Individual Differences*. 21, 215 – 222.

Papi, M. (2010). The L2 motivational self system, L2 anxiety, and motivated behavior: a structural equation modeling approach. *System*. 38. 467 – 479.

Phakiti, A. (2006). Modeling cognitive and metacognitive strategies and their relationships to EFL reading test performance. *Melbourne Papers in Language Testing*. 1. 53 – 95

Phakiti, A. (2008). Strategic competence as a fourth-order factor model: a structural equation modeling approach. *Language Assessment Quarterly*, 5, 1. 20 – 42

Purcell, E.T. (1983). Models of pronunciation accuracy. In Oller, J.W. (ed.) *Issues in language testing research*, Rowley, MA: Newbury House, 133–53.

Purpura, J.E. (1996). Modeling the relationships between test takers' reported cognitive and metacognitive strategy use and performance on language tests. Unpublished Ph.D. dissertation, University of California, Los Angeles.

Quintana, S. M., & Maxwell, S. E. (1999). Implications of recent developments in structural equation modeling for counseling psychology. *Counseling Psychologist, 27*, 486-527.

Raykov, T. & Marcoulides, G. A. (2006). *A first course in structural equation modeling (2nd ed.)*. Lawrence Erlbaum Associates, Publishers.

Rivers, D. J. (2011). Japanese national identification and English language learning processes. *International Journal of Intercultural Relations*. 35. 111-123

Sasaki, M. (1993). Relationships among second language proficiency, foreign language aptitude and intelligence: a structural equation modeling approach. *Language Learning* 43, 313–44.

Satorra, A., & Bentler, P. (1994). Corrections to test statistics and standard errors on covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis* (pp. 399-419). Thousand Oaks, CA: Sage.

Sawaki, Y., Stricker, L. J., & Oranje, A. H. (2009). Factor structure of the TOEFL Internet-Based test. *Language Testing*, 26(1), 005- 030.

Schoonen, R. (2005). Generalizability of writing scores: an application of structural equation modeling. *Language Testing*. 22 (1). 1-30

Schumacker, R. E, & Lomax, R. G. (2004). A beginner's guide to structural equation modeling (2nd ed). Lawrence Erlbaum Associates, Inc. Publishers.

Schreiber, J. D. (2008). Core reporting practices in structural equation modeling. *Research in Social and Administrative Pharmacy*. 4. 83 – 97.

Schroeders, U., Wilhelm, O., & Bucholtz, N. (2010). Reading, listening, and viewing comprehension in English as a foreign language: One or more constructs? *Intelligence*, 38. 562 – 573.

Shin, S. (2005). Did they take the same test? Examinee language proficiency and the structure of language tests. *Language Testing 22* , 1, 31-57.

MJAL5:3 Spring2013 ISSN 0974-8741

Shiotsu, T. & Weir, C. J. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing*. 24 (1), 99 - 128

Simon, H. (1953). "Causal ordering and identifiability". In Hood, W.C.; Koopmans, T.C.. *Studies in Econometric Method*. New York: Wiley. pp. 49–74.

Song, M. (2008). Do divisible subskills exist in L2(L2) comprehension? A structural equation modeling approach. *Language Testing 25*, 4, 435-464.

Steiger, J. H. (1990). Structural model evaluation and modification: An internal estimation approach. *Multivariate Behavioral Research*, *25,* 173-180.

Steiger, J. H., & Lind, J. C. (1980, May). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Swinton, S.S. & Powers, D.E. (1980). *Factor analysis of the TOEFL*. TOEFL Research Report 6. Princeton, NJ: Educational Testing Service.

Wang, L. S. (1988). A comparative analysis of cognitive achievement and psychological orientation among language minority groups: a LISREL approach. Ph.D. dissertation, University of Illinois, Urbana.

Weston, R. & Gore, P. A. (2007). A brief guide to structural equation modeling. The Counseling Psychologist, vol, 34, no. 5. 719 – 751.

Wright, S. S. (1921)."Correlation and causation". *Journal of Agricultural Research*. 20: 557–85.