Application of Structural Equation Modeling in EFL Testing: A Report of Two Iranian Studies

MOHAMMAD ALAVI
Department of English Language and Literature, Faculty of Foreign Languages, University of Tehran, Tehran, Iran

HAMED GHAEMI
Department of English Language and Literature, Faculty of Foreign Languages, University of Tehran, Kish International Campus, Kish, Iran

Bio Data:
Dr. S. Mohammad Alavi, is an Associate Professor of Applied Linguistics at English Department, Faculty of Foreign Languages, University of Tehran. He is the Vice Dean of Educational Affairs at Faculty of Foreign Languages and Literatures as well.

Hamed Ghaemi is a PhD candidate in TEFL at University of Tehran, International Campus, Iran. He is a TEFL instructor at Islamic Azad University. His areas of research interests are Language Testing, E-assessment, ESP and Teacher Development. He is currently the Head of IELTS department at the Jahan Elm Higher Education Institute, Mashhad, Iran.

Abstract
Structural equation modeling (SEM) is a statistical method used for testing and estimating causal relations using a combination of statistical data and qualitative causal assumptions. This paper reports two studies done by the researchers through the application of SEM. In the first study the relationship between EFL learners' affective constructs and their English achievement was assessed using SEM. Since both Structural coefficient and Goodness of fit are high, we conclude that not only is the model accepted but also the relationship between EFL learners' affective constructs and their English achievement is statistically significant. The second study investigated the relationship between Affective constructs and Study Process of EFL learners through the application of SEM. In this study the Goodness of Fit Index (GFI) turned out to be 0.99. Again like the previous study, since the Goodness of Fit Index (GFI) and structural coefficient are high, we conclude that there is a statistically strong relationship between affective constructs and study process of EFL learners.

Keywords: Structural Equation Modeling (SEM), English achievement, affective constructs, study process

1. Introduction

Structural Equation Modeling is a very powerful multivariate analysis method that includes particular versions of a number of other analysis techniques as special cases. The old definition of SEM was expressed by the geneticist Sewall Wright (1921), and officially defined by Judea Pearl (2000) using counterfactuals. The structural equation modeling (SEM) process focuses around two phases: validating the measurement model and fitting the structural model. The former is done mainly through confirmatory factor analysis, while the latter is carried out principally through path analysis with latent variables. Structural equation models can do both confirmatory and exploratory modeling, meaning that they are suitable for both theory testing and theory development. Confirmatory modeling mostly begins with a hypothesis that is usually presented in a causal model. The model is tested against the obtained data to determine how well the model fits the data (Bollen, and Long, 1993).

SEM can be used by identifying an analogous model and using data to estimate the values of free parameters. Frequently, the original hypothesis needs adjustment in light of model confirmation. Wright (1921) stated that when SEM is used purely for exploration, this is usually in the framework of exploratory factor analysis as in psychometric design (Wright, 1921).

Bollen, and Long (1993) pointed out that among the strengths of SEM is the ability to construct latent variables: variables which are not measured directly, but are estimated in the model from several measured variables each of which is predicted to 'tap into' the latent variables (Bollen, and Long, 1993). The qualitative causal assumptions are represented by the missing variables in each equation, and fading covariance among some error terms. These theories are testable in experimental studies and must be confirmed critically in observational studies (Gardner, Lalonde and Pierson, 1983).

2. Background to SEM

2.1. Model Specification

When used as confirmatory technique, the model created in SEM must be specified properly based on the type of analysis that the researcher is going to confirm (Anderson, J.C. and Gerbing, D.W. 1988). When building the proper model, the researcher uses two different types of variables, i.e. exogenous and endogenous variables. The difference between these two kinds of variables is whether the variable regresses on another variable or not. Exogenous variables can be documented as the variables sending out arrows, showing which variable it is predicting. Endogenous variables are perceived as the receivers of an arrow in the model (Anderson, J.C. and Gerbing, D.W. 1988).

As Austin and Calderon (1996) state, two main elements of models are differentiated in SEM: the first one is *structural model* depicting potential causal dependencies between endogenous and exogenous variables, and the second component is called *measurement model* showing the relations between latent variables and their indicators (Austin and Calderon, 1996).

In identifying pathways in a model, the researcher can speculate two types of relationships: the first kind of relationship is called *free* pathways, in which hypothesized causal relationships between variables are tested, and are left 'free' to vary, and the second type of relationship is the relationships between variables that already have an estimated relationship, usually based on previous studies, which are 'fixed' in the model (Austin and Calderon, 1996).

Bentler (1986) emphasized that a researcher can specify a set of theoretically possible models to appraise whether the proposed model is the best of the potential models. Not only should the researcher consider the theoretical rationales for building the model as it is, but the researcher must also take into account the number of parameters that the model must estimate to identify the model. An identified model is a model in which a particular parameter value exclusively identifies the model, and no other corresponding formulation can be given by a different parameter value (Bentler, 1986). If the number of data points is fewer than the number of estimated parameters, the final model would be "unidentified", because there are too few reference points to be responsible for all the variance in the model. However, the solution to this problem is to confine one of the paths to zero, which means that it is no longer part of the model (Bentler and Chou, 1987).

2.2. Estimation of Free Parameters

Bentler and Chou (1987) believe that parameter estimation is accomplished by evaluating the definite covariance matrices showing the relationships between variables and the estimated covariance matrices of the best fitting model. This is obtained through numerical maximization of a *fit criterion* as provided by maximum likelihood estimation accomplished by using a specialized SEM analysis program of which several exist (Bentler, P.M. and Chou, C.-P. 1987).

2.3. Assessment of Fit

Sasaki (1993) states that assessment of fit is a fundamental mission in SEM modeling: This includes forming the foundation for accepting or rejecting models and accepting one competing model over another. The output of SEM programs, such as LISREL, includes matrices of the estimated relationships between variables in the model. Assessment of fit basically calculates how similar the predicted data are to matrices including the relationships in the genuine data (Sasaki, 1993). Austin and Calderon (1996) indicate that every parameter of the model can be scrutinized within the estimated model to investigate how well the proposed model fits the theory (Austin, J.T. and Calderon, R.F., 1996).

The tests of SEM model are based on the assumption that the correct and relevant data have been modeled. The discussion of fit, in the SEM model, has led to a range of various recommendations on the precise use of the different fit indices and hypothesis tests.

Bollen (1989) states that Measures of fit are different in a number of ways. Traditional approaches to modeling start from a null hypothesis, however, the more modern approaches, such as AIC, focus on how little the fitted values diverge from a saturated model. Since different measures of fit use diverse components of the fit of the model, it is appropriate to report a selection of different fit measures (Bollen, K.A. 1989).

Some of the more frequently used measures of fit include:

2.3.1. Chi-Square. It is a function of the sample size and the difference between the observed covariance matrix and the model covariance matrix.

2.3.2. Root Mean Square Error of Approximation (RMSEA). Good models are considered to have a RMSEA of .05 or less. Models whose RMSEA is .1 or more have a poor fit.

For each of these measures of fit, a conclusion as to what shows a good-enough fit between the data and the model should represent other contextual factors such as sample size, the ratio of indicators to factors, and the overall complexity of the model (Bollen, K.A. 1989).

2.4. Model Modification

Bollen and Long (1993) emphasize that the model should be modified to improve the measures of fit, in that way estimating the most probable relationships between variables can be obtained (Bollen and Long, 1993). Nowadays, many programs can provide such modification indices. Besides improvements in model fit, it is essential that the modifications can also make theoretical logic (Sasaki, 1993).

2.5. Sample Size

In cases that SEM is the basis for a research hypothesis, the rules requiring the choosing of 10 observations per indicator in setting a lower bound for the adequacy of sample sizes have been extensively used (Nunnally, 1967). In one study Westland (2010) found that sample sizes in a special case of SEM literature averaged only 50% of the minimum required to draw the conclusions the studies claimed. On the whole, 80% of the research articles in the study drew conclusions from inadequate samples (Westland, 2010). Intricacies, which increase information demands in structural Equation modeling, increase with the number of probable combinations of latent variables. Sample size in SEM can be computed through two techniques: in the first one sample size is considered as a function of the ratio of indicator variables to latent variables, and in the second technique it is regarded as a function of minimum effect, power and significance (Nunnally, 1967).

3. The First Study

In the first study the relationship between EFL learners' affective constructs and their English achievement was assessed. Thus, the following null hypothesis was developed:

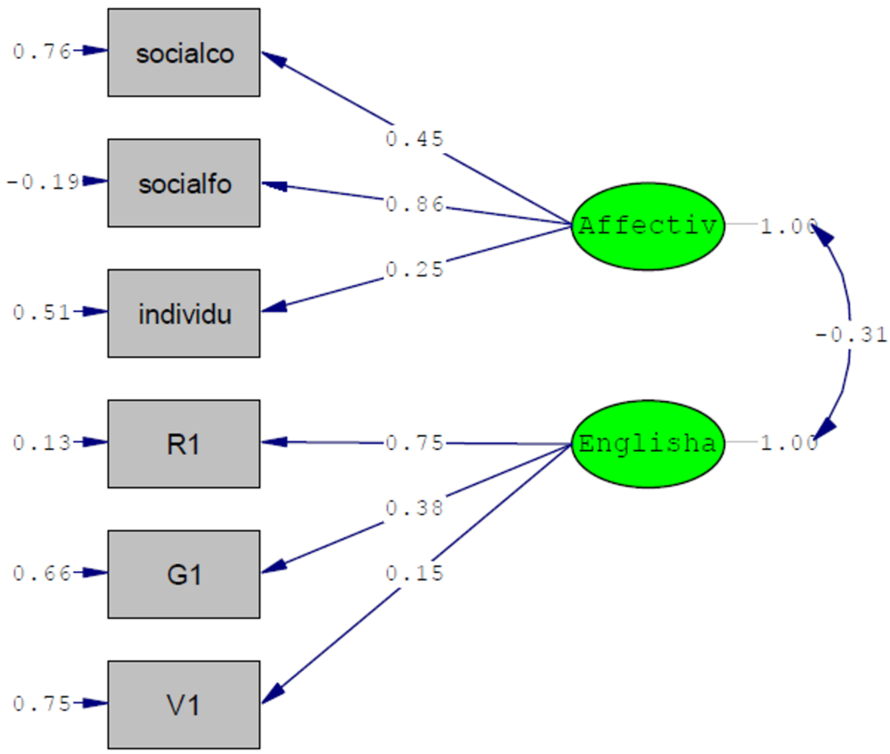
H0: There is a no meaningful relationship between EFL learners' affective constructs and their English achievement.

At the beginning of the study the questionnaire of Affective Constructs was developed and piloted on 36 intermediate EFL learners who were studying English at Jahan Elm Institute of Higher Education in Mashhad, Iran. The Cronbach's alpha was calculated and turned out to be 0.71. In order to assess the English achievement of the participants they were asked to participate in a MCHE exam. The test was held in the Jahan Elm Institute of Higher Education in Mashhad, Iran.

The questionnaire consisted of four main components, i.e. Social Composite, Social Focused, Individual Composite and Individual Focused. The MCHE exam also includes three sub-tests of Grammar, Vocabulary, and Reading. The data including the questionnaire and MCHE test were initially analyzed through SPSS Software then they were further analyzed through LISREL.

Having imported the data from SPSS software into LISREL and doing all the essential and required analysis, the following model was obtained.

Figure 1. Relationship between latent variables and observed variables



Chi-Square=8.47, df=8, P-value=0.38925, RMSEA=0.042

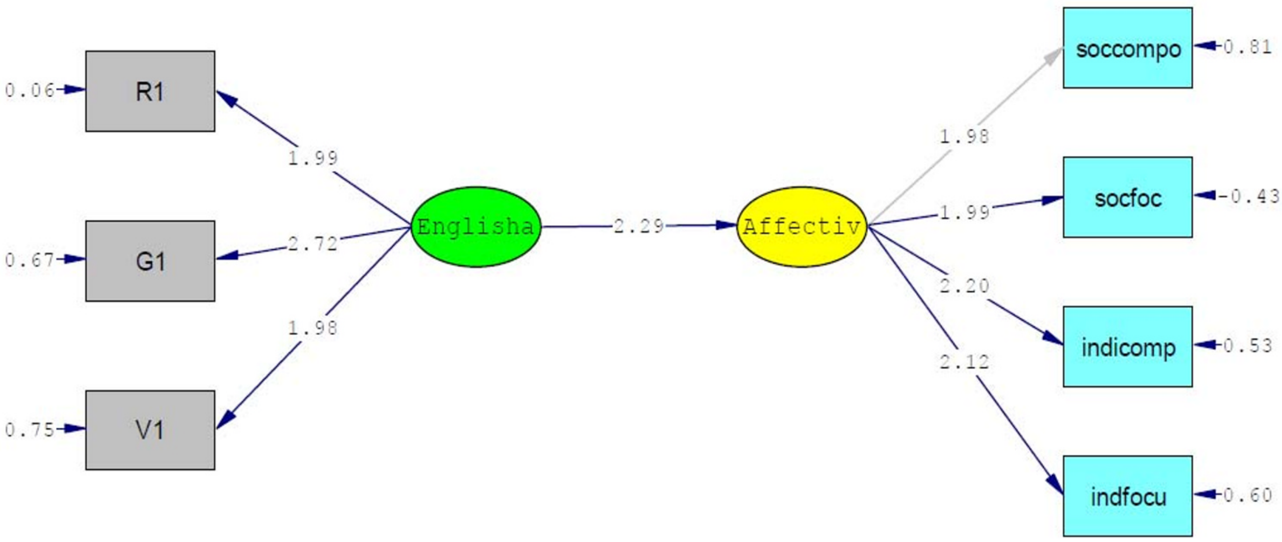
Since the Chi – Square equals 8.47 and the p-value is larger than 0.05, then we can draw conclusion that the model fits the data, i.e. the model is appropriate. This model only shows the relationship between the latent variables and the observed variables of the study.

The values which are written on each arrow are demonstrated in the *Estimated Mood*, and they cannot be appropriately interpreted. In all SEM models run in LSREL software, the values of *Estimated Mood* are not interpretable because there is no principle to which one can compare these values. In order to make the values interpretable, we should change the mood from *Estimated Mood* to *T-Value Mood*. Having changed the mood to T-Value, we see that all the values written on the arrows of the above model changed and are higher than 1.96 (1.96 is a predetermined principle value to which all the values are to be compared, when the

Critical Ratio (CR) is > 1.96 for a regression weight, that path is significant at the .05 level, i.e. its estimated path parameter is significant) (Ullman, 2001). As a result, we can conclude that there is a meaningful relationship between the observed variables (socialcomposit, socialfocused, and individualcomposite) and their latent variable, i.e. Affective Construct. Also, there is a meaningful relationship between Reading, Grammar and Vocabulary as the observed variables and English Achievement as the latent variable.

However, the main aim of the study was to assess the relationship between the Affective Constructs and English Achievement. That's why we need to analyze the relationship between these two variables as well. As figure 2 shows, there was a statistically strong relationship between Affective constructs and English Achievement of EFL learners, i.e. T-Value or Path Coefficient is 2.29 (Figure 2).

Figure 2. Relationship between English Achievement and Affective Constructs



As can be seen in Figure 2, the path coefficient between English Achievement and Affective Constructs is reported to be 2.29. Based on the SEM literature, if the T-value is larger than 1.96 we can conclude that there is a statistically significant relationship between the variables. Consequently, the null hypothesis developed for the study is rejected, that is, there is a meaningful relationship between English Achievement and Affective Constructs of the EFL learners.

As stated in the literature, SEM allows three significant jobs. Firstly, it estimates the Covariance Matrix. Table 1 below shows the Correlation matrix for the Observed variables of the first study. Second, SEM does the parameter specification including LAMBDA-X, PHI and THETA-DELTA. Parameter specification also calculates the Squared Multiple Correlations for the Variables of the study.

Table 1
Correlation Matrix of Observed Variables (First Study)

	Social Composite	Social Focused	Individual Composite	R1	V1	G1
Social Composite	0.96					
Social Focused	0.44	0.55				
Individual	0.22	0.22	0.57			
Composite	0.24	0.21	0.11	0.69		
R1	0.12	0.04	0.19	0.29	0.80	
V1	0.09	0.12	0.02	0.09	0.33	0.77
G1						

The next step is to estimate the Goodness of fit Statistics. Goodness of fit tests determines if the model being tested should be accepted or rejected. If the model is accepted, the researcher will continue to read the path coefficients in the model (Ullman, 2001).

Ullman (2001) states that a "good fit" is not the same as strength of relationship. One might have ideal fit when all variables in the model were entirely uncorrelated, provided that the researcher does not instruct the SEM software to constrain the variances. In reality, the lower the correlations predetermined in the model, the easier it is to find "good fit." The stronger the correlations, the more power SEM has to identify an incorrect model. When correlations are low, the researcher cannot reject the model at hand (Ullman, 2001).

When the variables have low correlation, the structural (path) coefficients will be low also. Researchers should report not only goodness-of-fit measures but also should report the structural coefficients so that the strength of paths in the model can be measured. In the case of the first study, the structural coefficient depicts a strong relationship between the variables of the study, i.e. EFL learners' affective constructs and their English achievement.

Goodness of Fit includes many parameters which are necessary for interpreting the results of the study. For the first step the Degree of freedom should be calculated. Here, for this study, it is estimated to be 8. Minimum Fit Function Chi-square is 9.35 when p-value is 0.31. The second parameter is called Estimated non-centrality parameter (NCP) that in this case it is estimated to be 0.47 and 90 percent Confidence interval for NCP is calculated as 0.0 ; 11.78. The next but the most important factor is Goodness of Fit Index (GFI). Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the model in question. In the case of this study, GFI is reported to be 0.92 (See Table 2). Since both Structural coefficient and Goodness of fit are high, we can conclude that not only is the model accepted but also the relationship between EFL learners' affective constructs and their English achievement is statistically significant.

Table 2
Parameters of Goodness of Fit (First Study)

	Minimum Fit Function Chi-square	Degree of freedom	Non-centrality parameter (NCP)	Goodness of Fit Index (GFI)
Value	9.35*	8	0.47	.92

* P= 0.31

4. The Second Study

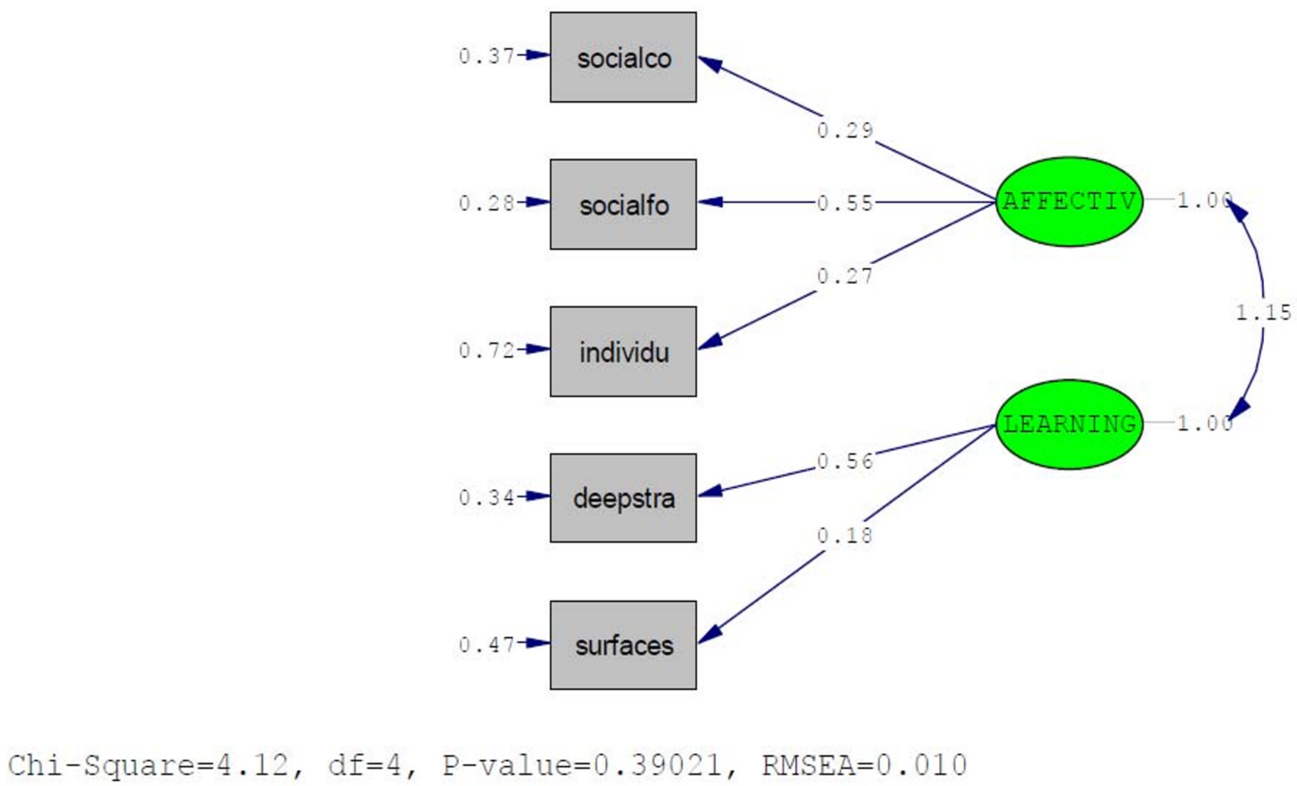
The second study investigated the relationship between Affective constructs and Study Process of EFL learners through the application of SEM. For the purpose of this study the following null hypothesis has been formulated:

H0: There is no relationship between EFL learners' Affective Constructs and their Study Process.

The questionnaire of Affective Constructs, developed and piloted in the previous study, was given to 88 EFL learners to complete it. The questionnaire consisted of four main components, i.e. Social Composite, Social Focused, Individual Composite and Individual Focused. The data regarding the second variable of the study, i.e. Learning Approach, was collected through the Learning Process Questionnaire (LPQ) developed by Biggs (1970). The LPQ is a 70-item self-report questionnaire that yields scores on two basic approaches to learning that is formed by Deep Strategies and Surface Strategies.

The data including the two questionnaires were initially analyzed through SPSS Software then they were further analyzed through LISREL. Having imported the data from SPSS software into LISREL and doing all the essential and required analysis, the following model was obtained.

Figure 3. Relationship between Latent Variables and Observed Variables

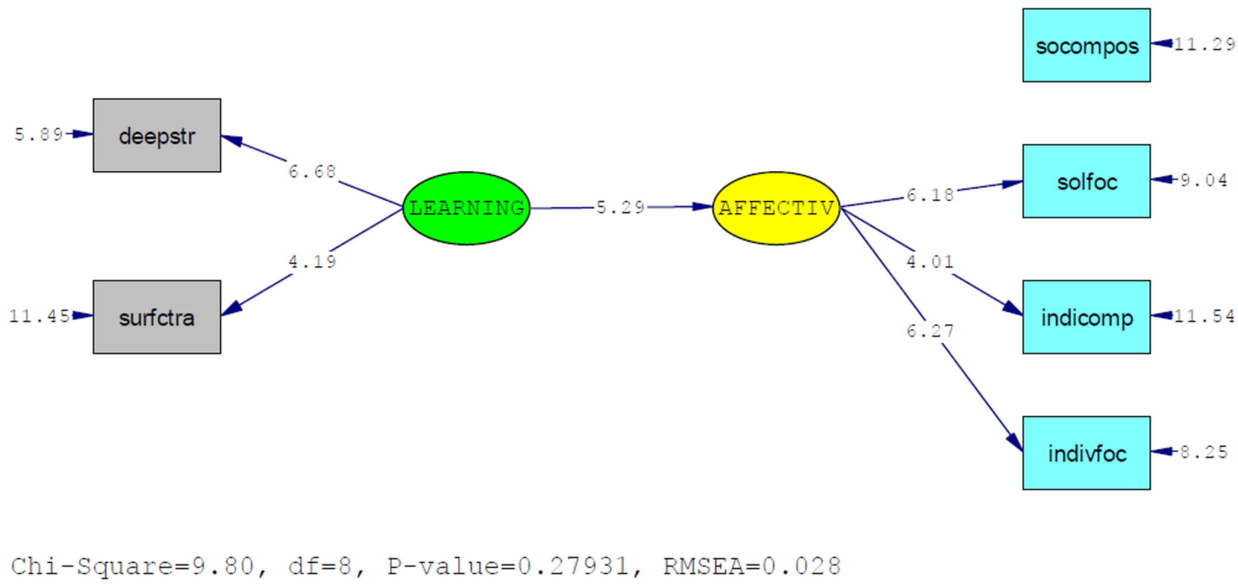


This model only shows the relationship between the latent variables and the observed variables of the study. As can be seen in figure 3, the Chi-Square is 4.12, and since the p-value is larger than 0.05, then we can conclude that the model fits the data, i.e. the model is appropriate. RMSEA equals 0.01, and this indicator also shows that the model is fit, since if the RMSEA is less than or equals to 0.05, we can come to the conclusion that the model fits the data and is appropriate.

The values written on each arrow is demonstrated in the *Estimated Mood*, therefore, as said earlier, they are not interpretable. In all SEM models written in LSREL software, the values of *Estimated Mood* are not interpretable because there is no principle to which one can compare these values. In order to make the values interpretable, we should change the mood from *Estimated Mood* to *T-Value Mood*. Having changed the mood to T-Value mood, we see that all the values written on the arrows of the above model are higher than 1.96 (1.96 is a predetermined, principle value to which all the values are to be compared). As a result, we can conclude that there is a meaningful relationship between the observed variables (socialcomposit, socialfocused, and individualcomposite) and their latent variable, i.e. Affective Construct except in the case of Social Composite. The T-Value is not significant in this case, so this latent variable is excluded from the further analysis. Also, there is a meaningful relationship between Deep Strategies and Surface Strategies as the observed variables and Learning Process as the latent variable.

Figure 3 depicts the relationship between the latent and observed variables of the study. However, the relationship between the two variables of the study, i.e. Affective Constructs and Learning Process, is shown in figure 4.

Figure 4. Relationship between Learning Approach and Affective Constructs



As it is indicated in Figure 4, the path coefficient or the T-value between Learning Process variables and Affective Constructs is reported to be 5.29, i.e. > 1.96. Therefore, the null hypothesis developed for the study is rejected as there is a meaningful relationship between Learning Process and Affective Constructs of EFL learners.

Table 3 below shows the Correlation matrix for the Observed variables of the second study. SEM also does the parameter specification including LAMBDA-X, PHI and THETA-DELTA. The Maximum Likelihood for all the observed and latent variables is estimated as well. The error variances for the variables of the study are negligible, however, the structural coefficient between each pair of variables are reported to be strong. The correlation matrix for the independent variables is also calculated. The correlation matrix computes the correlation coefficients of the columns of a matrix.

Table 3
Correlation Matrix of Observed Variables (Second Study)

	Social Composite	Social Focused	Individual Composite	Deep strategies	Surface Strategies
Social Composite	0.46				
Social Focused	0.18	0.59			
IndividualComposite	0.09	0.13	0.79		
Deepstrategies	0.18	0.37	0.17	0.65	
Surfacestrategies	0.07	0.10	0.11	0.10	0.50

In this study the Degrees of Freedom is estimated to be 4 and Minimum Fit Function Chi-Square equals 4.05 (P = 0.40). Goodness of Fit Index (GFI) was also calculated and turned out to be 0.99 (See table 4). Again like the previous study, since the Goodness of Fit Index (GFI) and structural coefficient are high, we can conclude that there was a strong relationship between affective constructs and study process of EFL learners.

Table 4
Parameters of Goodness of Fit (Second Study)

	Minimum Fit Function Chi-square	Degree of freedom	Non-centrality parameter (NCP)	Goodness of Fit Index (GFI)
Value	4.05*	4	0.49	.90

* P=0.40

5. Discussion and Conclusion

A search for applications of SEM in the field of language assessment in the context of Iran will undoubtedly not turn up more than a few papers at most. This low level of interest in SEM among Iranian language testing researchers is probably due to many reasons, the most significant ones are the lack of a pedagogic introduction to SEM for language testing research, very few instances of SEM application to language assessment data, and very little discussion of the virtues and the restrictions of SEM for the field of language assessment.

SEM applications are so extensive today that Marcoulides and Schumacker (1996) utter that

> the use of the term structural equation modeling is broadly defined to accommodate models that include latent variables, measurement errors in both dependent and independent latent constructs, multiple indicators, reciprocal causation, simultaneity and interdependence (p. 1).

In the 1980s, Gardner and other second language acquisition researchers employed SEM with the data obtained from SLA researches (Gardner, Lalonde and Pierson, 1983; Gardner et al., 1987; Gardner, 1988; Clement and Kruidenier, 1985; Ely, 1986) to scrutinize motivation and attitude as parameters that influence second language acquisition. The most current SEM applications in language assessment include Sasaki (1993), who investigated the relationships among second language proficiency, foreign language aptitude, and intelligence, Kunnan (1995), who explored the influence of some test taker characteristics on test performance in tests of English as a foreign language, Purpura (1996), who examined 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.

As said earlier, the application of SEM in language testing in Iran doesn't have remarkable background. The current study aimed at reviewing the application of SEM in EFL testing in the context of Iran. In the first study, the relationship between Affective constructs and English Achievement of EFL learners was assessed through the application of SEM. The path coefficient between English Achievement and Affective Constructs is reported to be 2.29. Accordingly, the null hypothesis developed for the study is rejected, that is, there is a meaningful relationship between English Achievement and Affective Constructs of the EFL learners.

For this study, the Degree of Freedom is estimated to be 8. Minimum Fit Function Chi-square is 9.35 when p-value is 0.31. Estimated non-centrality parameter (NCP) is estimated to be 0.47 and 90 percent Confidence interval for NCP is calculated as 0.0; 11.78. The next but the most important factor is Goodness of Fit Index (GFI). Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the model in question. In the case of this study, GFI is reported to be 0.92. Since both Structural coefficient and Goodness of fit are high, we can conclude that not only is the model accepted but also the relationship between EFL learners' affective constructs and their English achievement is statistically significant.

The second study investigates the relationship between Affective Constructs and Study Process of EFL learners using SEM. The path coefficient or the T-value between Learning Process variables and Affective Constructs is reported to be 5.29, i.e. > 1.96. Thus, the null hypothesis formulated for the study is rejected as there is a meaningful relationship between Learning Process and Affective Constructs of EFL learners.

In this study the Degrees of Freedom is estimated to be 4 and Minimum Fit Function Chi-Square equals 4.05 (P = 0.40). Goodness of Fit Index (GFI) was also calculated and turned out to be 0.99. Again like the previous study, since the

Goodness of Fit Index (GFI) and structural coefficient are high, we can conclude that there was a strong relationship between affective constructs and study process of EFL learners.

Finally, due to the many scientifically - reported advantages of SEM, such as: latent growth modeling, multilevel SEM models, and approaches for dealing with missing data and with violations of normality assumptions, application of SEM in the field of language assessment is highly recommended.

References

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103,* 411–23.

Austin, J. T., & Caldero´n, R. F. (1996). Theoretical and technical contributions to SEM: An updated bibliography. *Structural Equation Modeling, 3,* 105–75.

Bentler, P. M. (1986). Structural modeling and Psychometrika: An historical perspective on growth and achievements. *Psychometrika, 51,* 35–51.

Bentler, P. M., & Chou, C. P. (1987). Practical issues in SEM. *Sociological Methods and Research, 16,* 78–117.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

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

Clement, R., & 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.

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

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

Kunna, A. J. (1995). *Test taker characteristics and test performance: A structural modeling approach*. Cambridge: Cambridge University Press.

Marcoulides, G. A., & Schumacker, R. E. (1996). *Advanced SEM: Issues and techniques*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Nunnally, J. C. (1967). *Psychometric theory.* New York: McGraw-Hill.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press: Cambridge.

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

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

Simon, H. (1953). Causal ordering and identifiability. In Hood, W. C., & Koopmans, T. C. (Eds.), *Studies in Econometric Method*. New York: Wiley.

Ullman, J. B. (2001). Structural equation modelling. In Tabachnick, B. G., & Fidell, L. S. (Eds.), *Using multivariate statistics* (pp. 653- 771). Needham Heights, MA: Allyn & Bacon.

Westland, J. C. (2010). Lower bounds on sample size in structural equation modeling. *Electron. Comm. Res. Appl, 9*(6), 476–487. doi:10.1016/j.elerap.2010.07.003

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