COST PERFORMANCE PREDICTION FOR INTERNATIONAL CONSTRUCTION PROJECTS USING MULTIPLE REGRESSION ANALYSIS AND STRUCTURAL EQUATION MODEL: A COMPARATIVE STUDY

D.Y. Kim, S.H. Han, H. Kim, and H. Park[1]
[1] Department of Civil & Environmental Engineering, Yonsei University, Seoul, Korea

Abstract

Overseas construction projects tend to be more complex than domestic projects, being exposed to more external risks, such as politics, economy, society, and culture, as well as more internal risks from the project itself. It is crucial to have an early understanding of the project condition, in order to be well prepared in various phases of the project. This study compares a structural equation model and multiple regression analysis, in their capacity to predict cost performance of international construction projects. The structural equation model shows a more accurate prediction of cost performance than does regression analysis, due to its intrinsic capability of considering various cost factors in a systematic way.

Keywords: Forecasting, International Construction Projects, Multiple Regression Analysis, and Structural Equation Model

1. Introduction

According to the *Engineering News Record* [1], the total annual volume of the global construction market is estimated to reach around $ 4.6 trillion, and approximately $ 300 billion of this market (7% of the total volume) is available to foreign contractors. Given the rapid trend of globalization, the overseas construction market that will be accessible to foreign contractors is expected to expand to up to $ 600 billion per year within a decade [1]. However, the profitability of overseas construction projects has a tendency to be lower than that of domestic projects, because foreign construction works are affected by more diverse external risks, such as politics, economy, society, and culture, as well as more noteworthy internal risks [2]. To successfully conduct an international construction project, it is important to have an early understanding of the project performance by assessing profitability, cost structure, and construction duration in its early stage.

A variety of research has been conducted to produce methodologies that predict project performances and selection of a promising project [3, 4, 5, and 6]. These research intend to improve international contractors' revenue by helping them formulate contract strategies and acquire promising overseas projects. This study aims to suggest a model that predicts cost performance of a project by utilizing a structural equation model and multiple regression analysis. The structural equation model (SEM) complements the multiple regression analysis by considering various project factors in a systematic and organized manner. The paper further provides a theory explanations and research conclusions through a comparative analysis between the multiple regression analysis and structural equation model.

2. Literature Review

Modeling of complicated prediction processes in construction management has been attempted through many types of analysis methods [5, 7, 8, 9, 10, and 16]. These prediction models are selected to fit each subject of interest, considering the merits/demerits of the models and the research characteristics, including research objectives, intention to use the model, and the scope of data required.

Typically, statistical methods can show a clear causality of the prediction results to ensure the result in the form of statistically reliable figures [2]. Among those statistical methods, multiple regression analysis is one of the most widely used for modeling and forecasting [7], because it requires a relatively simple process [2, 8, 9, 10, 11, and 12]. However, the modeling method using multiple regression analysis has the flaw of ignoring all the potential measurement error of the observed variables [13]. On the other hand, the SEM is superior to multiple regression methods in taking into consideration the measurement error.

Since the SEM was developed by Jöresjig in the 1970s, the use of this model has expanded rapidly, along with the development of computer science from the 1980s to the 1990s. The reasons why SEM has recently been favored are: first, it can identify a genuine causal relationship between an independent variable and dependent variable by taking the measurement error of the observed variable into account; second, modeling a concept that is difficult to directly measure or explicitly quantify is possible using a latent variable; and third, it can model indirect effects as well as the direct causal relationship among variables [13]. In addition, it is possible to visualize complex mutual relations among variables through a graphical representation. A modeling through SEM has been constantly increasing in the construction management area, dealing with complex phenomena and relationships. For example, Mohamed [5] used SEM to model the joint venture (J/V) performance of overseas construction projects. Molenaar et. al. [7] suggested using SEM to predict the possibility of disputes during the project. In addition, Islam and Faniran [14] modeled the project planning effects, such as cost and schedule, through SEM, and Wang and Cheung [15] also proposed a SEM-based prediction model as to whether partnering could be successfully implemented.

SEM can perform substantial modeling of the relationship of variables through a path diagram; thereby it is presumed that SEM is suitable for overseas construction projects that are usually confounded by complex mutual relations of diverse variables. Based on the result of our precedent research, this study aims to model the relationship among various influencing factors and project cost performance of overseas construction through SEM, in order to produce management strategies and to select promising overseas construction projects by examining the project cost performance in advance.

3. Structural Equation Model

3.1 Review of relevant factors

Overseas construction is vulnerable to various political, economical, social, and cultural factors of the host country [12, 16, 17, 18, 19, and 20]. We reviewed the influencing factors that were presented in preceding research, and identified 64 influencing factors, which were set up into 5 groups based on the experts' opinions regarding the characteristics of international projects. We also investigated the cost performance of 126 international projects conducted by Korean contractors in the last decade, in relation to the 64 influencing factors; this is to examine the relationship between the influencing factors and the performance of the overseas construction projects. The precedent research [2] presented a model to predict the causal relationship between the 64 factors and project profitability through multiple regression analysis; however, limitation of a regression model was recognized, leading to the identification of future research to conduct multilateral analysis of the relation of these influencing factors through the structural equation method.

3.2 Description of Structural Equation Model

SEM is a systematic combination of confirmatory factor analysis, multiple regression analysis, and path analysis. SEM consists of a measurement model and structural model: the former incorporates confirmatory factor analysis, while the latter reflects multiple regression analysis and path analysis [13]. SEM includes a latent variable and observed variable based on the execution of observation: the former refers to a hypothetical concept, which cannot be observed nor measured directly; the latter is a variable that can be measured directly, so it plays a vital role in measuring the latent variable ultimately by interconnecting with it. The SEM variables are also classified into an exogenous variable and endogenous variable, based on whether they influence or get influenced. The former refers to a kind of independent variable, which influences other variables; the latter refers to the one influenced by other variables, directly or indirectly. A typical SEM is represented in the forms of exogenous latent variable, endogenous latent variable, exogenous observed variable, and endogenous observed variable by compounding them with one another. More importantly, SEM has an error variable that considers both of measurement error and structural error to reflect the actual phenomena more accurately.

For instance, as Figure 1 shows, latent variables include 'home background' and 'academic performance' as conceptual variables which are marked in an oval, while the observed variables include 'income levels', 'education levels', 'language grades', and 'mathematics grades,' which are marked in a quadrangle. Namely, 'home background' and 'academic performance', which bear latent meaning, cannot be measured directly, so that 'income levels' and 'parents' education levels' are used to indirectly measure 'home background', and similarly 'language grades' and 'mathematics grades' are used to measure 'academic performance' as well. The errors related to the observed variables found during the measurement process are measurement error variables, and the errors related to 'academic performance' are structural error variables, respectively.

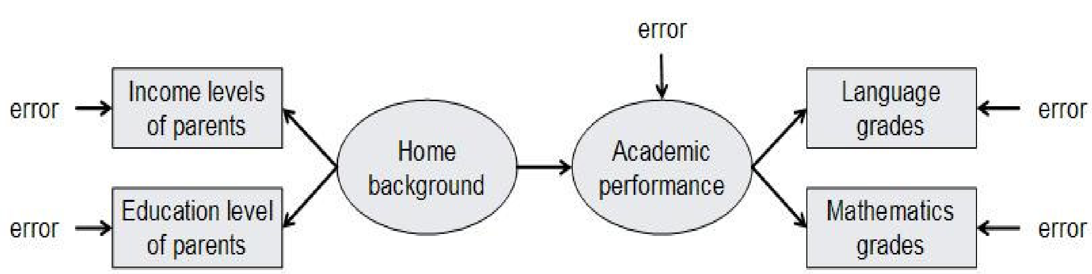


Figure 1: Example of SEM – Home Background and Academic Performance [13]

3.3 Project Cost Performance Model

In order to structure the causal relationship between the 64 performance-influencing factors and the project performance of overseas construction found in our precedent study, this study hypothesized eight latent variables affecting performance of overseas projects, as shown in Figure 2, by investigating the existing literature related to influencing factors in overseas construction.

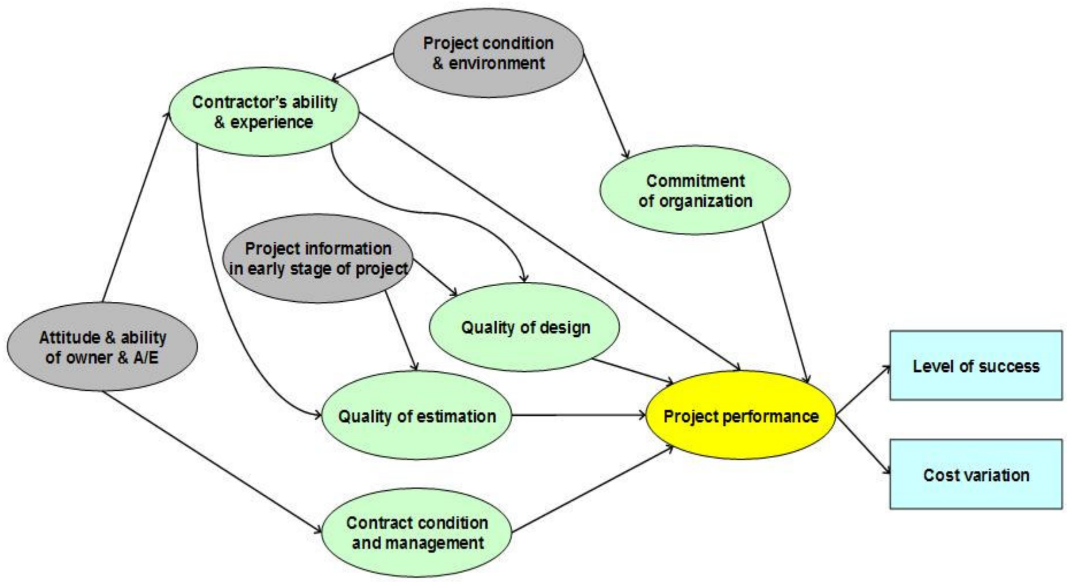


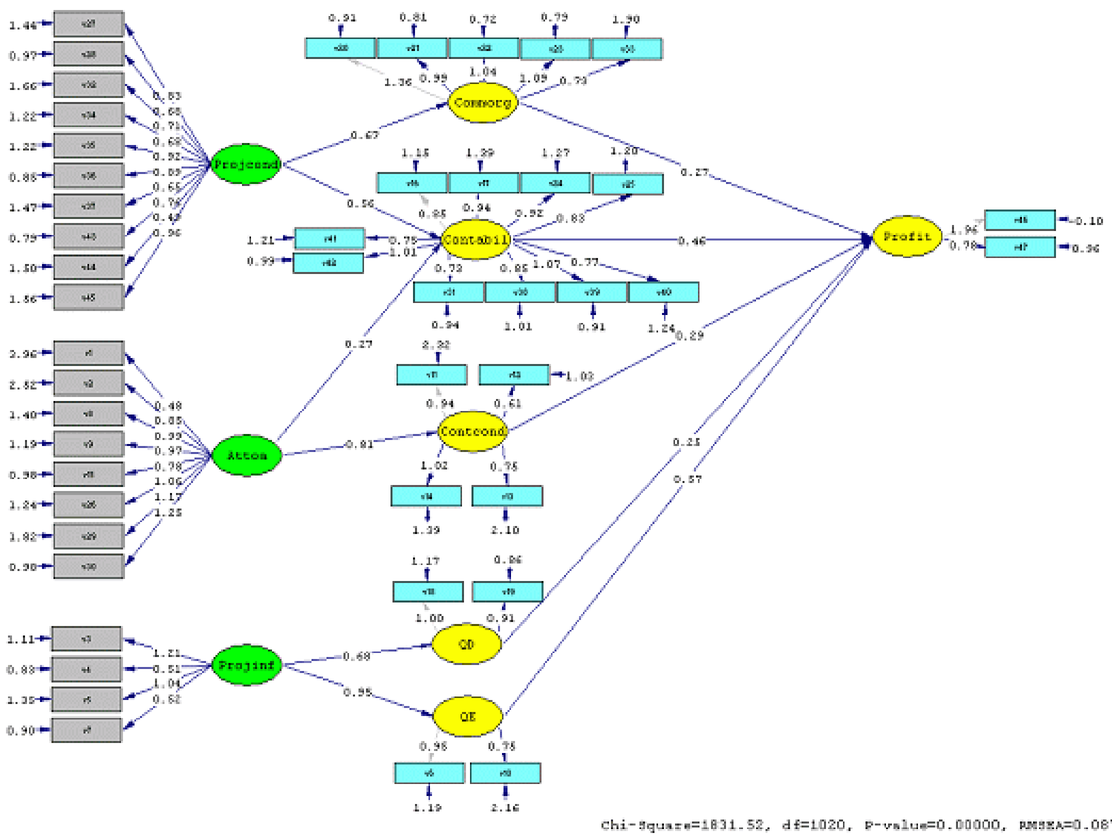
Figure 2: Project Cost Performance Model

In establishing the cost performance model, we adopted 'project environment & conditions,' 'attitudes of owner and A/E (Architecture and Engineer),' and 'quality of project information collected at the early stage of the project' as exogenous latent variables. This is very relevant to the findings of the existing literatures that include the environment of the host country, owner's characteristics, and early project information as boundary conditions for overseas construction projects [4, 16, and 18]. It is presumed that endogenous latent variables consist of five elements, including 'organization system and commitment of workers', 'contractor's capability and experience', 'contract condition and management skill', 'quality of design', and 'quality of estimate'. All of these 8 variables were adopted to conform to major variables affecting performance in the multiple regression model through exploratory factor analysis in our precedent study, and the 64 performance-influencing factors were arranged in consideration of the meaning of each factor through the confirmatory factor analysis of the 8 latent variables. Finally, the logical relationship of the mutual effects among variables was made in a path diagram, as shown in Figure 2.

4. Analysis of Structural Equation Model

This study established a structural equation model by utilizing 126 real project data collected during our precedent study which was conducted to identify the relationship between project performance and performance-influencing factors. This study used the

LISREL (LInear Structural RELation) program for analysis of a structural equation model. LISREL is widely used as the representative program of SEM. This study drew an analysis result, as displayed in Figure 3 by going through repeated amendments to produce a model suitable to data collected through the SEM model-generating method.



Note; ellips: latent variable, rectangle: observed variable, arrow: direction of relationship
'Projcond': project condition & environments, 'Attoe': attitude of owner & A/E, 'Commorg': commitment of organization, 'Contabil': contractor’s ability & experience, 'Contcond': contract condition, 'QD': quality of design, 'QE': quality of estimation

Figure 3: Analysis Results of SEM

The final model reveals that the project performance is directly affected by 5 endogenous latent variables (Figure 3). The five endogenous latent variables, including 'organization system and commitment of workers,' 'contractor's capability and experience,' 'contract condition and management skill,' 'quality of design,' and 'quality of estimate,' are the factors that directly affect the project performance, while the remaining three exogenous latent variables, including 'project environment & conditions,' 'attitudes of owner and A/E (Architecture and Engineer),' and 'quality of project information collected at the early stage of the project,' affect the project performance indirectly.

The one-way relation of latent variables is presented as a causal relationship in the regression model, and the cost performance or profitability, which is a dependent variable, can be presented in the multiple regression model through five endogenous latent variables. According to the analysis, the multiple regression model with five endogenous latent variables has a R-square of 0.71. This implies that 126 sample data can explain the cost performance to the degree of 71% through the SEM. Considering the 65% of the multiple regression model in our precedent study, SEM's R-square has risen by taking the structural relation of variables into account. Furthermore, even though SEM came up with 64 influencing factors like the multiple regression model, it can visualize the relationship of variables by simply identifying the structural relation of factors and variables. The model is more suitable to the project data for reflecting error terms. For example, the advanced information about the host country has a regression coefficient of 1.21 for the exogenous latent variable 'Projinf', which means that this factor is the highest among the observed variables in influencing the pertinent latent variable, and the exogenous latent variable 'Projinf' has an influence of 0.95 on the endogenous latent variable 'QE'. Collectively, these factors, along with other factors, affect the endogenous latent variable 'QE', which exercises an influence of 0.57 on the project performance.

On the other hand, one of the indices for SEM is the goodness of fit index (GFI). It explains how well data fits to the model, generally ranging from 0 to 1.0. GFI is interpreted similarly with R^2 of the regression analysis, and is rarely affected by a change in sample size and violation of multivariate normality; therefore it is an index clearly explaining goodness of fit of the suggested model. Since the universally recommended level is above 0.8 [21], it can be noted that the model reflects the data properly with a GFI of 0.872 (Table 1).

Table 1: Analysis Result of SEM

Description	Results
Number of sample moments	126
Degrees of freedom	1,020
Chi-Square at model convergence	1,831.29
RMSEA	0.087
Goodness of Fit Index (GFI)	0.872
Adjusted Goodness of Fit Index (AGFI)	0.785
Parsimony Goodness of Fit Index (PGFI)	0.741

5. Comparison of Two Models

Analysis results from multiple regression analysis in the precedent research are shown in table 2. The results of multiple regression analysis suggested a linear relationship between risk factors and project cost performance. From the results, practitioners can recognize important factors influencing project cost performance. For example, ‘quality of estimation’ and ‘project information in the early stage of a project’ have a large degree of influence which records a regression coefficient of 0.619 and 0.527, respectively. However, in international construction projects, project performance cannot be adequately explained just through the linear relations, because they have the possibility of having more complex problems such as the interrelation between risk factors and observed errors. Consequently, it can be more helpful for understanding performance algorithms by using SEM because we can see the relationship between risk factors visually and systematically.

Table 2: The Result of Multiple Regression Analysis [2]

Composition (Group Factors)	Unstandardized coefficient		t	Significance
	B	Standard error		
Constant	4.670	0.138	33.873	0.000
F10 Quality of estimation	0.619	0.139	4.467	0.000
F7 Project information in early stage of a project	0.527	0.139	3.800	0.000
F4 Attitude and ability of owner & A/E	0.506	0.139	3.647	0.000
F12 Contract conditions and management	0.499	0.139	3.596	0.001
F2 Project conditions – resource delivery, labor skill, etc	0.477	0.139	3.440	0.001
F5 Commitment of organization – PM competency, etc	0.470	0.139	3.393	0.001
F1 Contractor's ability & experience	0.402	0.139	2.899	0.005
F8 Quality of design	0.379	0.139	2.734	0.008

Meanwhile, although the 126 projects used to build the forecasting model have many similarities with the other projects, it is noted that they are just particular samples of a population. Therefore, the models needed to be validated. Toward this end, another 15 projects were selected to validate the models by comparing the output from models to the actual success level of performance. A T-test was used to find the differences between the previous 126 projects and the additional 15 projects. As a result of the T-test, the level of significance of the two-tailed test is 0.962, which means that there are few differences between the two distributions of the project groups. The deviation between output from the two models and actual performance levels were compared, and the accuracy of each model was expressed by means of a percentage (Table 3). As for the multiple regression model, the average deviance of level of performance is 0.82 and the overall accuracy is 86.3%. Similarly, those from the SEM are 0.49 and 91.8%, respectively. Although both models are worthwhile in predicting the scale-based performance for any probable overseas projects in the early state of project initiation, it is pointed out that the SEM is more accurate and powerful in recognizing the complex structures of variables and their underpinned causes-and-effects relationships.

Table 3: Validation of Two Models

Project #	Performance level (Response)	Regression model (1~7)		SEM (1~7)	
		Forecasting result	Deviation	Forecasting result	Deviation
1	7	6.71	0.29	6.48	0.52
2	5	4.18	0.82	4.91	0.09
3	5	3.28	1.72	4.28	0.72
4	5	7.00	2.00	6.89	1.89
5	7	6.15	0.85	6.91	0.09
6	7	7.00	0.00	6.49	0.51
7	1	1.00	0.00	0.98	0.02
8	5	2.96	2.04	4.79	0.21
9	5	4.83	0.17	4.63	0.37
10	6	6.28	0.28	6.95	0.95
11	4	5.98	1.98	5.36	1.36
12	6	7.00	1.00	6.08	0.08
13	1	1.00	0.00	1.25	0.25
14	1	1.37	0.37	1.15	0.15
15	2	2.83	0.83	2.10	0.10
Average			0.82		0.49
Overall Accuracy		86.3%		91.8%	

6. Conclusion

This study produced a model to predict the cost performance of overseas construction projects, using structural equation model (SEM). SEM is a way to systematically combine confirmatory factor analysis, multiple regression analysis, and path analysis, in order to have a holistic view of a complex interdependent phenomenon. The data extracted from 126 construction projects were used to develop the SEM, and additional 15 construction project data were used to test the model. The comparison analysis between SEM and multiple regression analysis indicated that SEM has a strong potential to accurately and reliably predict the cost performance of international construction projects. Future research is required to develop a more refined SEM to better explain the causal relationships of diverse cost influencing factors. This will lead to more accurate cost prediction of international construction projects.

Acknowledgment

The authors wish to thank the Korea Research Foundation for providing the funding that made this study possible.

References

[1] **Engineering News Records (ENR). (2005).** "The top 225 international contractors". *Engineering News Records*, December, 2005, 24-40.

[2] **Han, S.H. and, Kim, D.Y. (2006).** "Risk-based Profit Prediction Model for International Construction Projects". *Journal of Civil Engineering, KSCE*, 26(4-D), 635-647.

[3] **Chua, D.K.H., Li, D.Z., and Chan, W.T. (2001).** "Case-based Reasoning Approach in Bid Decision Making". *J., Constr. Engrg. and Mgmt., ASCE*, 127(1), 35-45.

[4] **Han S.H., and Diekmann, J.E. (2001).** "Approaches for Making Risk-based Go/No-go Decision for International Projects". *J., Constr. Engrg. and Mgmt., ASCE*, 127(4), 300-308.

[5] **Mohamed S. (2003).** "Performance in International Construction Joint Ventures: Model Perspective". *J., Constr. Engrg. and Mgmt., ASCE*, 129(6), 619-626.

[6] **Dikmen, I., and Birgonul, M.T. (2004).** "Neural Network Model to Support International Market Entry Decisions". *J., Constr. Engrg. and Mgmt., ASCE*, 130(1), 59-66.

[7] **Molenaar, K., Washington, S., and Diekmann, J. (2000).** "Structural Equation Model of Construction Contract Dispute Potential". *J., Constr. Engrg. and Mgmt., ASCE*, 126(4), 268-277.

[8] **Russell, J.S., and Jaselskis, E.J. (1992).** "Predicting construction contract failure prior to contract award". *J., Constr. Engrg. and Mgmt., ASCE*, 118(4), 791-811.

[9] **Sanders, S.R., and Thomas, H.R. (1993).** "Masonry productivity forecasting model". *J., Constr. Engrg. and Mgmt., ASCE*, 119(1), 163-179.

[10] **Diekmann, J.E., and Girard, M.J. (1995).** "Are contract disputes predictable?". *J., Constr. Engrg. and Mgmt., ASCE*, 121(4), 355-363.

[11] **Molenaar, K.R., and Songer, A.D. (1998).** "Model for public sector design-build project selection". *J., Constr. Engrg. and Mgmt., ASCE*, 124(6), 467-479.

[12] **Chan, A.P.C., Ho, D.C.K., and Tam, C.M. (2001).** "Design and Build Project Success Factors: Multivariate Analysis". *J., Constr. Engrg. and Mgmt., ASCE*, 127(2), 93-100.
[13] **Bae, B.R. (2005).** *LISREL Structural Equation Model : Understanding & Application*, Chunglam Publishing Company. Korea.
[14] **Islam, M.D.M., and Faniran, O.O. (2005).** "Structural Equation Model of Project Planning Effectiveness". *Construction Management and Economics*, 23, 215-223.
[15] **Wang, P.S.P., and Cheung, S.O. (2005).** "Structural Equation Model of Trust and Partnering Success". *J., Mgmt. in Engrg., ASCE*, 21(2), 70-80.
[16] **Pinto, J.K., and Mantel, S.J. Jr. (1990).** "The Causes of Project Failure". *IEEE Transactions on Engineering Management*, 37(4), 269-276.
[17] **Zhi, H. (1995).** "Risk Management for Overseas Construction Projects". *International Journal of Project Management*, 13(4), 231-237.
[18] **Bing, L., and Tiong, L.K. (1999).** "Risk Management Model for International Construction Joint Ventures". *J., Constr. Engrg. and Mgmt., ASCE*, 125(5), 377-384.
[19] **Wang, S.Q., Tiong, R.L.K., Ting, S.K., and Ashley, D. (2000).** "Evaluation and Management of Political Risks in China's BOT Projects". *J., Constr. Engrg. and Mgmt., ASCE*, 126(3), 242-250.
[20] **Baloi, D., and Price, A.D.F. (2003).** "Modeling Global Risk Factors Affecting Construction Cost Performance". *International Journal of Project Management*, 21(4), 261-269.
[21] **Kline, R. B. (2005).** *Principles and Practice of Structural Equation Modeling*, 2nd Ed., The Guilford Publications, Inc., New York.