Journal of Positive School Psychology http://journalppw.com
2023, Vol. 7, No. 10, 77-100

Analytic Choice Between And Within Multiple Regression And Structural Equation Modeling Approaches

Abdulmajeed A Alharbi

Department of Psychology, College of Education, Imam Abdulrahman Bin Faisal University, Saudi Arabia.

Abstract :This research demonstrated whether the analytic choice of using a multiple regression or structural equation modeling methodology affected the results of faculty research productivity in Saudi Arabia. This study not only showed the differences between multiple regression and structural equation modeling results but also the disparity of results within each type of analysis. The results indicated that using either a multiple regression or structural equation modeling methodology delivered different results in terms of significant predictors and the model's overall explained variance. Further, differential outcomes produced by the various structural equation modeling models employed illustrate how the incorrect specification of formative (i.e., causal) indicators can result in worse data-fitting models. Implications for selecting analytic procedures are discussed.

Keywords: Multiple Regression, Structural Equation Modeling, Formative Measurement Model, Reflective Measurement Model, Analytic Procedures.

Introduction

The importance of data analysis lies in the fact that it is a procedure through which outcomes can be extracted from an examination (Silberzahn et al., 2018). Researchers have claimed that while statistical analysis procedures consist of noteworthy components, such as hypotheses, the strong relationship between outcomes and the selected logical technique is often neglected (Silberzahn et al., 2018). Nayak and Hazra (2011) asserted that choosing a suitable analytic procedure ought not to be done randomly during the phase of analyzing the collected data but rather, when designing the study. Indeed, the research question has to guide the choice of statistical methodology. Emphasizing the importance of the relationship between the statistical analysis and the results of the study, Simundic (2016) declared that the legitimacy of the outcomes and inferences is contingent upon the nature of the chosen analytic techniques. Thus, the important aspect of the study is to use the survey data collected to explore whether there is a disparity in results when applying two different statistical analyses (i.e., multiple regression [MR] and structural equation modeling [SEM]) to data collected on faculty research productivity.

Both MR and SEM are appropriate statistical methods when examining comparative structural models, and they can also be utilized to estimate the direct and indirect effects of variables in a model (Musil et al., 1998). While SEM is increasingly being used in many fields, multiple researchers consider it a complicated technique (Nachtigal et al., 2003); therefore, they are using MR instead for its ease of application (Li, 2011). MR is also utilized because it is an applicable method of analysis when facing challenges in constructing measurement models (Li, 2011). The debate regarding interpreting results from MR and advanced analytic techniques such as SEM has been decades long (Grace & Bollen, 2005). However, Musil et al.

(1998) argued that SEM is a superior statistical method compared to MR and path analysis; for example, unlike MR, SEM employs a maximum-likelihood method that generates unbiased values because it does not assume independent error terms. Furthermore, SEM is a favorable statistical analysis method that not only considers measurement errors but also simultaneously structures various criterion variables, two functions that are not incorporated in MR, Analysis of variance (ANOVA), path analysis, or nested models (Wang & Wang, 2012). Petter et al. (2007) recommended utilizing the most effective statistical procedure that best matches the data, rather than using an analysis technique because it is straightforward.

Research Questions and Hypotheses

The researcher addressed whether the analytic choice of using either MR or SEM affects the results of the research question. The primary question is, Is there variation between the analysis results of MR and SEM when investigating the factors affecting faculty research productivity at Saudi universities? To address this question, the following secondary questions need to be considered:

1. To what extent does varying the order of entering variables into hierarchical MR models produce different results?
2. To what extent does using different MR modeling approaches (i.e., standard and statistical [stepwise]) affect the results?
3. To what extent does using various SEM models affect the results?

By examining the variation in these analyses, this study clarified to what extent similarities and differences exist between the results of MR and SEM analyses. The researcher hypothesizes that there will be a disparity in the outcomes analyzed by using MR and SEM techniques.

Literature Review

Multiple Regression

MR has become increasingly popular among researchers because of its simplicity (Li, 2011). Regression methods (i.e., standard, statistical, and hierarchical) aim to depict the amount of variation in the dependent variables accounted for by the predictor variables (Meyers et al., 2017). In a statistical (i.e., stepwise) regression analysis, the variable that contributes to the largest increase in R^2 is retained, and the nonsignificant variables are removed from the model (Meyers et al., 2017). The statistical method is suitable when the focus of the study is to determine the strongest predictors of the dependent variables and to eliminate the nonsignificant variables in order to build a statistical model (Meyers et al., 2017; Tabachnick et al., 2007). However, Tabachnick et al. (2007) recommended cautiously interpreting the conclusions of the statistical method because it requires a large and very representative sample. Because researchers are increasingly aware of the drawbacks of statistical regression, such as external validity issues, its use has been decreasing (Meyers et al., 2017); thus, it is gradually becoming globally perceived as an unrecommended analysis (Ruengvirayudh & Brooks, 2016).

Regardless of its popularity and ease of use, standard regression is a suitable method when the intent of the research is model building by testing all independent variables simultaneously in terms of their significant contribution to variability in the criterion variable (Meyers et al., 2017). However, because standard regression examines all variables simultaneously, it should only be used when the purpose of the study is to investigate whether the variation in the criterion variable was significantly predicted from the set of all predictor variables (Meyers et al., 2017).

Still, by only relying on standard regression, researchers are unable to resolve some research questions (Meyers et al., 2017). Thus, when the question is to examine whether each set of predictor variables entered into blocks, after controlling for other variables, significantly predicts the criterion variables, a hierarchical regression method should be considered instead. The variables in the hierarchical method are controlled differently from those in the standard method, resulting in different information regarding the relationship between the independent and dependent variables; thus, varying the order of entering variables into the models should be carefully based on a theoretical rationale (Tabachnick et al., 2007; Warner, 2013). One possible issue often overlooked when using hierarchical regression is researchers' exclusive use of the change in R^2 when testing the overall model (Ludlow & Klein, 2014), particularly considering that adding multiple predictor variables in a single block can sometimes cause R^2 to be inflated (Meyers et al., 2017). Indeed, sometimes an overall model includes variables that did not significantly predict the dependent variable in previous models, thus casting doubt on the actual relationship between predictors and the criterion variable. Therefore, the existence of suppressor variables should be examined (Ludlow & Klein, 2014; Pedhazur, 1997); otherwise, researchers risk making inappropriate conclusions about the interrelationship among variables in the model.

Structural Equation Modeling

Although the chosen statistical method has an effect on the analysis results, many analysts do not consider this fact (Silberzahn et al., 2018). Musil et al. (1998) asserted, "Depending on the nature of the hypothesized model, regression analysis, SEM, or another statistical technique may be the procedure of choice" (p. 280). Also, although both MR and SEM can be conducted to test theory-based models, SEM was chosen for further analysis because of its advantages over MR (Musil et al., 1998). For example, measurement models allow SEM to examine the measurement errors as well as estimate them in the latent variable (Musil et al., 1998). Furthermore, SEM possesses techniques capable of delivering unbiased parameter estimates and thereby more reliable results (Wang & Wang, 2012). While both MR and SEM can be utilized as statistical methods to test the hypothesized model, SEM further examines the measurement error by employing measurement models (Musil et al., 1998). These measurement models are reflective and formative indicators (Jarvis et al., 2003).

Because of the Type I error inflation that occurred when testing multiple hypotheses about the dependent variables, SEM might be a more suitable tool. Thus, this study utilized SEM because of its advantage over MR in analyzing both observed and latent variables simultaneously, which MR cannot do (Kline, 2016; Musil et al., 1998). While MR assumes no measurement error associated with causal indicators, that is not practically supported (Diamantopoulos et al., 2008; Kline, 2016), SEM (here, using a formative measurement model) estimates measurement error in the latent variable, called disturbance (Kline, 2016). Using formative indicators is rarely practiced in measurement models (Diamantopoulos et al., 2008). Also, Petter et al. (2007) stated that it is obvious that "formative constructs have been specified incorrectly as reflective even in publications that have appeared in premier scholarly journals" (p. 644). Therefore, when choosing formative over reflective indicators, researchers should consider that measured variables in formative measurement models are neither interchangeable nor limited to having a pattern of covariances (Coltman et al., 2008; Jarvis et al., 2003; Kline, 2006). Moreover, a reflective measurement model should be avoided when the variables served as predictors of the latent variable.

The multiple indicators and multiple causes (MIMIC) model can be considered when a measurement model contains a latent variable (or construct) that consists of its formative and reflective indicators (Diamantopoulos, 2011; Franke et al., 2008; Kline, 2006). Wang et al. (2015) asserted the effectiveness of formative measurement models by saying that “the ability to validate formative measurement has increased in importance as it is used to develop and test theoretical models” (p. 83). When a latent variable is caused by the predictors in the formative measurement model (Kline, 2006), this latent variable measured by reflective indicators can be distinctly elucidated because it is a “common factor accounting for the covariance among its outcomes” (Howell, 2014, p. 144).

Methodology

Study Design and Sampling Procedures

This research investigated whether the analytic choice between MR and SEM affects the results when examining the factors affecting research productivity in Saudi Arabia. The researcher used SEM to investigate whether personal, professional, and institutional variables are related to faculty research productivity on a number of measures. Thus, Brown (2015) stated that “CFA [confirmatory factor analysis] is used to verify the number of underlying dimensions of the instrument (factors) and the pattern of item-factor relationships (factor loadings)” (p. 1). Also, SEM will be utilized in this study because of its advantages over MR analysis in that SEM possesses techniques that can provide unbiased parameter estimates and thus more accurate results (Wang & Wang, 2012).

First, MR will be performed to examine whether personal, professional, and institutional variables predict the faculty research productivity in Saudi Arabia. The researcher hypothesizes that there is a relationship among these three factors and the research productivity in Saudi Arabia. Thus, the researcher will use AMOS version 26 and SPSS to address the research question: Is there variation between the analysis results of MR and SEM when investigating the factors affecting faculty research productivity at Saudi universities?

The study’s target population was all faculty members at four Saudi universities who worked as assistant professors, associate professors, or professors for the academic year of 2018–2019.

Data Sources

The researcher used an existing survey developed by Alzuman (2015), which was dedicated to the research performance of faculty at Saudi universities. Slight amendments have been made to the survey to eliminate qualitative questions that do not meet the purpose of the study. The survey contains three sections (i.e., personal, professional, and institutional information) with a total of 11 variables and 21 items. Five questions measured the personal variables of age, gender, marital status, and nationality; six questions measured the professional variables of academic rank and instructional (i.e., teaching) duties; and seven questions measured the institutional variables of financial support, research incentives, teaching/research assistants, research environment, and institutional and administration duties (i.e., workload). To obtain faculty responses on these items, the researcher asked the participants to respond to the questionnaire using two scales: (1) a 5-point Likert scale and (2) a 4-point Likert scale (1 = Not important at all, 2 = Not very important, 3 = Important, 4 = Very important). Reliability was evaluated for the 21 items that were measured by the three scales. The Cronbach’s alpha for scale 1, which applied the

5-point Likert scale, was .77. The Cronbach's alpha for scale 2, which asked whether or not specific services were provided by the university, was .66. Finally, the Cronbach's alpha for scale 3, which applied the 4-point Likert scale, was .77. Table 1 displays the mean (M), standard deviation (SD), and Cronbach's alpha for each scale.

Table 1 Number of items (N), Means, Standard deviations, and Cronbach' alpha for scales

	N	M	SD	Cronbach' alpha
Scale 1	6	15.35	4.018	.768
Scale 2	7	80.21	1.786	.664
Scale 3	8	27.78	3.429	.773

To measure the faculty research output (i.e., the dependent variable), the respondents were asked about their research productivity during the past five years (i.e., since the academic year 2014–2015). The items for this variable tend to gauge the number of their publications in refereed academic or professional journals, books, and book chapters as well as papers presented at scientific conferences and books edited or translated. They will indicate their research productivity by choosing from the following response options: 0 = Never published, 1 = Had published 1–2, 2 = Had published 3–4, 3 = Had published 5–6, and 4 = Had published over 6.

Procedures

The researcher contacted the deanship of the Research at each university selected to participate in the study. The emails that were sent to the deans included the purpose of the study and the request to send the surveys to the prospective participants. The data were collected through three parts of the survey. The first part contained items measuring the institutional variables, the second part focused on questions about the personal variables, and the third part aimed to gauge the professional variables.

Data Analysis

The researcher used AMOS version 26 and SPSS version 22.0 to conduct the MR and SEM to investigate whether variation exists between the statistical results obtained through MR and SEM when analyzing the factors affecting the faculty research productivity at Saudi universities.

The researcher utilized SEM to test the causal structure that shows the relationship among the personal, professional, and institutional variables and faculty research productivity. Initially, the researcher screened the data to address linearity, normality, influential cases, and multicollinearity. After examining the identification of the model, the researcher analyzed the effectiveness of the overall model using the most popular goodness-of-fit indices χ^2: CFI, RMSEA, and AIC; these statistics were then used to provide comparative results among the models analyzed in this study. CFI was chosen because it is one of the most informative statistics in terms of comparing models (Schermelleh-Engel et al., 2003), and a value close to or greater than .95 for CFI shows a good fit (e.g., Beauducel & Wittman, 2005; Brown, 2015; Meade, 2008; Meyer et al., 2017). Brown (2015) defined RMSEA as a fit index that evaluates lack of model fit and examines the degree to which a model fits well in the population. A value of .08 or less for the RMSEA index suggests a reasonable error of approximation (Browne & Cudeck, 1993). AIC was used because it is the best fit index to compare alternative models and smaller indexes indicate better fit (Kline, 2016).

Also, the researcher used Cronbach's coefficient alpha to test the reliability of measurement.

In addition to SEM, the researcher conducted a hierarchical MR to examine whether independent variables can predict the faculty research productivity at Saudi universities. The variables were entered into the analysis in three blocks to see whether the variation in the dependent variable is statistically explained by each indicator after controlling for other variables. Block 1 contained the personal variables (i.e., age, gender, marital status, and nationality); block 2, the professional variables (i.e., academic rank and instructional duties); and block 3, the institutional variables (i.e., financial support, research incentives, teaching/research assistants, research environment, and institutional duties). Within each model, the individual predictors were examined by analyzing the statistical significance of beta weights (β). Then, the overall fit of the regression model (R^2) was tested along with other related statistics. Also, the researcher used the descriptive statistics to explore the variables of the study. To verify the appropriateness of the statistical analysis used, the data were scanned to both identify the potential outliers and examine the regression assumptions (i.e., linearity, normality, collinearity, and homoscedasticity). Furthermore, the researcher tested the null hypothesis of MR and each independent variable and used the significance level of .05.

Results

This research was conducted to explore the degree of similarities and differences between the results of MR and SEM regarding the question of whether personal, professional, and institutional variables account for a significant amount of the variation in research productivity of faculty. The total size of this sample was 175 faculty members. Table 2 displays the demographic characteristics for this sample (i.e., gender, citizenship, and academic rank).

Table 2 Participant's Demographic characteristics, gender, citizenship, and rank

Variable	N	%
Gender		
Male	98	56.0
Female	76	43.4
Citizenship		
Saudi	118	67.4
NonSaudi	56	32.0
Academic rank		
Assistant professor	92	52.6
Associate professor	31	17.7
Professor	18	10.3

The following section presents the results of MR in analyzing whether personal, professional, and institutional variables predict the faculty research productivity in Saudi Arabia. Initially, the researcher replicated a previous study conducted by Alzuman (2015), which used MR analysis to investigate factors affecting faculty research productivity in Saudi Arabia, then applied two additional methods of entering models in the hierarchical regression analysis. Next, the researcher described the results of conducting three different models of MR analysis: hierarchical, standard (i.e., simultaneous), and statistical. Finally, the researcher demonstrated the results of performing the same analyses after compressing the data into fewer composite measures (i.e., predictor variables).

Multiple Regression (MR)

Q1. To what extent does varying the order of entering variables into the hierarchical MR models produce different results?

The hierarchical MR analysis consisted of comparing three different procedures for entering predictor variables into the model. In procedure 1, predictor variables were entered into the analysis in three blocks (or models): Block 1 contained the institutional variables; block 2, the personal variables; and block 3, the professional variables. This method of analysis directly replicates Alzuman's (2015) study. In procedure 2, the personal variables were entered into block 1; the professional variables, into block 2; and the institutional variables, into block 3. The rationale behind entering variables in this order was that the primary model (i.e., institutional variables) should be entered in the final model (or block) after controlling for other variables. In procedure 3, the professional variables were entered in block 1, the institutional variables, into block 2; and the personal variables into block 3.

The following section illustrates the results of the three procedures for entering predictor variables into the model and their relationship to the dependent variables—(a) publication in refereed and professional journals and (b) published books—using a listwise deletion method resulting in 98 valid cases out of a total of 175.

Publication in Refereed and Professional Journals

Entering personal, professional, and institutional variables into different blocks produced different results in analyzing whether these variables predict published articles in referred or professional journals (See Tables 3,4, and 5). For example, entering institutional variables in block 1 did not significantly predict published articles in refereed or professional journals, R^2 change = .307 $F(21, 97) = 1.604$, $p = .071$, but they did significantly explain variability when entered in both block 1 after controlling for professional variables, R^2 change = .307, $F(21, 97) = 1.729$, $p = .046$, and block 3 when controlling for personal and professional variables, R^2 change = .308, $F(21, 97) = 1.830$, $p = .033$. Entering institutional variables into different blocks produced different results regarding which individual variables, β, significantly predicted outcomes. When entering these variables into block 2, research centers and research assistants significantly predicted published articles in refereed or professional journals, $p = .030$ and $p < .001$, respectively. However, when entering institutional variables into block 3, contribution to theoretical development also significantly predicted publication, $p = .045$. Also, professional variables did not significantly predict published articles in block 1, $R^2 = .101$, $F(6, 97) = 1.703$, $p = .129$, nor did they lead to significant increases in explained variance when entered into block 2, R^2 change = .068, $F(6, 97) = 1.177$, $p = .326$, and block 3, R^2 change = .087 $F(6, 97) = 1.801$, $p = .113$. However, adding these variables into block 3 and controlling for institutional and personal variables yielded a significant overall model, R^2 =. 479, $F(32, 97) = 1.865$, $p = .017$.

Overall, the final models may include significant individual predictors that are distinct from those in earlier models across procedures. In procedure 1, block 2, which included institutional and personal variables, showed that course release time significantly predicted outcomes (t (97) = 2.189, p = .032) but that this variable was not statistically significant in the overall model when adding professional variables, (t (97) = -.974, p = .334) (As illustrates in Table 3). Also, in procedure 3, while contribution to theoretical development was a significant predictor in the third (overall) model, t (97) = 2.045, p = .045, it did not significantly predict published articles in block 2, t (97) = 1.809, p = .075 (As illustrates in Table 5).

Table 3 Results of Procedure 1 for publication in refereed or professional journals

	β*	R^2 change	Overall R^2	adj R^2	AIC
Model 1		.307	307	.116	93.234
Model 2		.085	. 392*	.169	90.425
Research assistant	-.443				
Research centers	-.297				
Course release time	.278				
Citizenship	.308				
Model 3		.087	. 479*	.222	87.351
Experience	.325				
Contribution to theoretical development	.266				
Research centers	-.418				
Research assistant	-.487				
Citizenship	.283				

* < .05

Table 4 Results of Procedure 2 for publication in refereed or professional journals

	β*	R^2 change	Overall R^2	adj R^2	AIC
Model 1		.102	102	. 054	86.604
Model 2		.068	. 171	. 064	90. 871
Model 3		.308*	. 479*	.222	87.351
Experience	.325				

Contribution to theoretical development	.266				
Research centers	-.418				
Research assistant	-.487				
Citizenship	.283				

* < .05

Table 5 Results of Procedure 3 for publication in refereed or professional journals

	β*	R^2 change	Overall R^2	adj R^2	AIC
Model 1		. 101	.101	. 042	88.770
Model 2		. 307*	. 408*	.180	89.813
Experience	.228				
Research centers	-.312				
Research assistant	-.485				
Model 3		.071	. 479*	.222	87.351
Experience	.325				
Theoretical development	.266				
Research centers	-.418				
Research assistant	-.487				
Citizenship	.283				

* < .05

Published Books

The results showed that varying the order of entering variables into the models across blocks produced different results for which individual predictors significantly predicted research productivity within those blocks (As illustrates in incentives was the only institutional variable that statistically predicted published books (Table 6). Additionally, procedure 2 showed that adding Tables 6, 7, and 8). In procedure 1, block 1 showed that the institutional variables (i.e., financial incentives, access to academic library, promotion system, and research assistants) significantly predicted publishing books, $p < .05$. However, in the overall model, financial professional variables to the personal variables in block 2 produced a significant β for time spent in administration or teaching work, $t(97) = -2.113$, p = .037, but that this variable was not significant when adding the institutional variables to the analysis, t (97) = -.952, p = .345 (Table 7). Procedure 2 also showed that age significantly predicted published books in block 2, t (97) = 2.787, p = .007, but not in blocks 1 and 3, t (97) = 1.711, p = .090 and t (97) = 1.802, p = .076, respectively.

Table 6 Results of Procedure 1 for published books

	β*	R^2 change	Overall R^2	adj R^2	AIC
Model 1		.409*	.409*	. 245	96.957
Promotion system	-.222				
Financial incentives	-.233				
Access to academic library	.291				
Research assistant	-.260				
Model 2		.082	. 490*	. 304	92.408
Financial incentives	-.231				
Access to academic library	.251				
Gender	.303				
Model 3		.024	. 514*	. 275	99.657
Financial incentives	-.248				
Gender	.293				

* < .05

Table 7 Results of Procedure 2 for published books

	β*	R^2 change	Overall R^2	adj R^2	AIC
Model 1		.248 *	.248 *	. 207	88.583
Gender	.500				
Model 2		.074	. 321*	. 234	90.474
Gender	.432				
Age	.378				
Teaching duties	-.213				
Model 3		.024	. 514*	. 275	99.657
Financial incentives	-.248				
Gender	.293				

* < .05

Because of the small sample size resulting from the previous analyses using a listwise deletion method, the researcher used a pairwise deletion method to assess whether it would provide different results. By using a pairwise deletion method, the total valid cases were increased to 175. The results showed that the pairwise method produced different results from those using a listwise deletion method for all three dependent variables. Table 9 shows the similarities and differences in results using these two deletion methods in terms of the explained shared variances, R^2, and its significance.

Table 8 Results of Procedure 3 for published books

	β*	R^2 change	Overall R^2	adj R^2	AIC
Model 1		. 117	.117	. 059	106.235
Model 2		. 315*		. 214	104.920
Financial incentives	.260				
Access to academic library	.262				
Research assistant	-.271				
Model 3		.082	. 514*	. 275	99.657
Financial incentives	-.248				
Gender	.293				

* < .05

Table 9 Overall regression for models of hierarchical multiple regression with original data (R^2 and p)

	Listwise deletion method	Pairwise deletion method
The average RP	.550*	.405
Publication in journals	.479*	.382
Published books	.514*	.368

* < .05

Q2. To what extent does using different MR modeling method approaches (i.e., standard and statistical [stepwise]) affect the results?

Standard Multiple Regression. Standard or simultaneous MR was performed. The predictive value of each independent variable was evaluated when controlling for other variables. Table 10 shows the overall regression analysis for each dependent variable. The results for explained shared variance and for which variables significantly predicted research output were identical to those in hierarchical regression models.

Table 10 Overall regression for standard multiple regression model with original data (R^2 and p)

Variables	R^2
Publication in journals	.479*
Published books	.514*
Published book chapters	.554*
Presenting papers	.378
Edited and translated books	.614*
Publication in journals	.479*

* < .05

Statistical Multiple Regression. Statistical (or data-driven regression) was executed utilizing two methods: stepwise and forward regressions. By default, the criterion values were .05 for F-to-enter and .10 for F-to-remove. Using a listwise deletion method, the results showed that the final model significantly predicted publication in refereed and professional journals, $R^2 = .165$, F(3, 97) = 6.179, p < .05, 11.302, p = .029).

The results showed that statistical (stepwise) MR analysis produced results different from those of the standard MR analysis regarding R^2 values and significant β values. Tables 11 and 12 display the differences between standard and statistical MR analyses.

Table 11 Overall regression for statistical and standard multiple regression models with original data (R^2, β, and p)

	Statistical regression	Standard regression
Publication in journals	.165*	.479*
Published books	.351*	.514*

* < .05

Table 12 Overall models (β) for statistical and standard multiple regression models with original data

	Statistical		Standard	
	β	p	β	p
Publication in journals				
Contribution to theoretical development	.223	.020	.266	.045

Citizenship	.203	.037	.283	.015
Research centers	------	------	-.418	.006
Research assistant	------	------	-.487	.000
Experience	.213	.029	.325	.027
Published books				
Financial incentives	-.191	.026	-.248	.037
Gender	.353	.000	.293	.021
Research assistant	-.232	.007	-.229	.071
Accessing to academic library	.201	.024	.110	.370

Because of the small sample size used in the original data set, which included 32 items, the data were compressed into 18 items to achieve more power. For institutional variables, the mean for each set of items having the same scale was calculated, which resulted in a set of three items. For personal variables, dummy coded variables were created for the items that measure marital status, children, age, and citizenship. For professional variables, one variable, academic rank, was coded as a dummy variable. By reducing the total number of predictor variables, the resulting data included 168 cases that had no missing data when using the listwise deletion method compared to 98 cases in the original data. Applying the same procedures that were used on the original data, the results of the compressed data showed that the two data sets using the same analyses yielded different results.

Hierarchical Multiple Regression. The R^2 resulting from the analyses of the compressed data for each dependent variable was much smaller than that resulting from the original data (As illustrates in Table 13). Furthermore, several predictor variables that were significantly related to each dependent variable in the original data reported no significant contributions with the compressed data (See Tables 14, 15, and 16). For example, institutional variables significantly predicted publication in refereed or professional journals when entered into block 1, R^2 change = .047, $F(3, 167) = 2.710$, $p = .047$. However, these variables did not significantly increase the explained variance in publishing refereed or professional journals when entered into block 2, R^2 change = .024, $F(3, 167) = 1.537$, $p = .207$, or block 3, R^2 change = .019, $F(3, 167) = 1.253$, $p = .293$. Thus, compressing institutional items to fewer variables produced different results regarding R^2 and significant β values. Also, the professional variables that significantly predicted publishing in journals were associate professor and professor, $p = .009$ and $p = .002$, respectively, which were not significant in analyses using the original data, $p = .187$. Therefore, coding dummy variables for rank variables in the compressed data produced significant changes in the ability to predict journal publications. Furthermore, treating marital status as a categorical variable versus a nominal variable, as in the original data, would make it a significant individual predictor in the overall models as well as models 1 and 2.

Table 13 Overall regression for models of hierarchical multiple regression with compressed and original data (R^2 and p)

	Compressed data (N=168) R^2	Original data (N = 98) R^2
Publication in journals	.266*	.479*
Published books	.275*	.514*
Published book chapters	.295*	.554*
Presenting papers	.160	.378
Edited and translated books	.358*	.614*

* < .05

Table 14 Results of Procedure 1 for publication in refereed or professional variables with compressed data

	β*	R^2 change	Overall R^2	adj R^2	AIC
Model 1		.047 *	.047 *	. 030	145.451
Mean Q1-6	.235				
Model 2		.131 *	. 178*	. 120	136.564
Marital status	-.304				
Children 1-2	.337				
Children 3-5	.444				
Citizenship	-.308				
Model 3		.088 *	. 266*	. 178	131.562
Marital status	-.311				
Citizenship	-.235				
Associate professor (Rank=2)	.226				
Professor (Rank=3)	.301				

* < .05

Table 15 Results of Procedure 2 for publication in refereed or professional variables with compressed data

	β*	R^2 change	Overall R^2	adj R^2	AIC

Model 1		.145 *	.145 *	. 102	137.312
Marital status	-.286				
Children 3-5	.426				
Citizenship	-.330				
Model 2		.103 *	. 248*	. 174	129.748
Marital status	-.303				
Citizenship	-.249				
Associate professor	.235				
Professor	.338				
Model 3		.019	. 266*	. 178	131.562
Marital status	-.311				
Citizenship	-.235				
Associate professor	.226				
Professor	.301				

* < .05

Table 16 Results of Procedure 3 for publication in refereed or professional variables with compressed data

	β*	R^2 change	Overall R^2	adj R^2	AIC
Model 1		. 169*	. 169*	. 133	130.479
Associate professor	.247				
Professor	.321				
Model 2		.024	. 193*	.141	131.617
Associate professor	.232				
Professor	.272				
Model 3		. 074	. 266*	. 178	131.562
Associate professor	.226				
Professor	.301				
Marital status	-.311				
Citizenship	-.235				

* < .05

Statistical Multiple Regression. The stepwise and forward analyses produced identical results, which were different than those in the original data. The original data reported higher R^2 than the compressed data, .381 > .229. Further, the predictor variables that significantly predicted the scores of the average research productivity in the compressed data were different than those in the original data. Table 17 displays the findings of the statistical MR analysis for the two data sets.

Table 17 Overall regression for statistical multiple regression models with original and compressed data sets (R^2 and p)

	Original data (N = 98)	Compressed data (N=168)
The average RP	.381*	.229*
Publication in journals	.165*	.197*
Published books	.351*	.243*

* < .05

From the previous analyses, it was clear that using different MR analyses with the original and compressed data sets produced different results in terms of the overall regression for the final model (R^2) and significant contributions of the individual predictors (β).

Q3. To what extent does using various SEM models affect the results?

There were two different SEM models used to examine the causal relationship among all predictor values and faculty research productivity, a latent factor measured by five previously described indicators, and those models were analyzed using both the original and compressed data sets.

Regarding the original data, model 1 included 32 causal predictors that affect faculty research productivity (Figure 1). Model 1 produced a good-fitting model, $\chi^2(175, 133) = 279.654$, CFI = .914, RMSEA = .080, AIC = 1493.654 (Table 18). This model was conducted with the option to correlate all predictor variables. However, this SEM model produced different results than those analyzed with MR in terms of R^2 and significant β. Contrary to MR, none of the predictor variables have a significant β with SEM analysis, p > .05, and the total research productivity, for example, has $R^2 = .516$ compared to $R^2 = .550$ with MR. It is hard to conclude that SEM or MR analyses provided greater R^2 values because no pattern was determined. Table 19 depicts the different R^2 values for MR and SEM analyses.

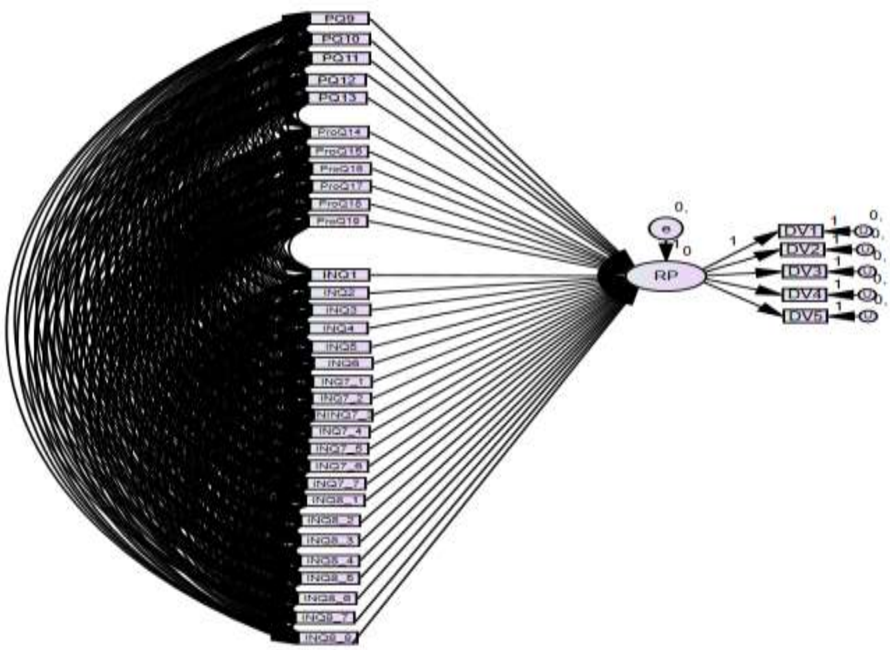

Figure 1. Model 1 with original data

Table 18 Model fit indices (χ^2, CFI, RMSEA), df, and AIC (N = 175)

Model	χ^2	df	CFI	RMSEA	AIC
Original data					
Model 1	279.654	133	.914	.080	1493.654
Compressed data					
Model 2	155.020	77	.939	.076	599.020

Table 19 Results of MR and SEM (R^2) for original data

Dependent variables	Standard MR	SEM
Publication in journals	.479	.014
Published books	.514	.728
Published book chapters	.554	.731
Presenting papers	.378	.126
Edited and translated books	.614	.906
Total research productivity	.550	.516

Regarding the compressed data, model 2 included 18 causal indicators that affect the faculty research productivity (Figure 2). Model 2 yielded a good-fitting model, $\chi^2(175, 77) = 155.020$, CFI = .939, RMSEA = .076, AIC = 599.020 (Table 18). Although model 2 produced a good-fitting model, it provided R^2 values that are different than those analyzed with the original data but also delivered nonsignificant β values.

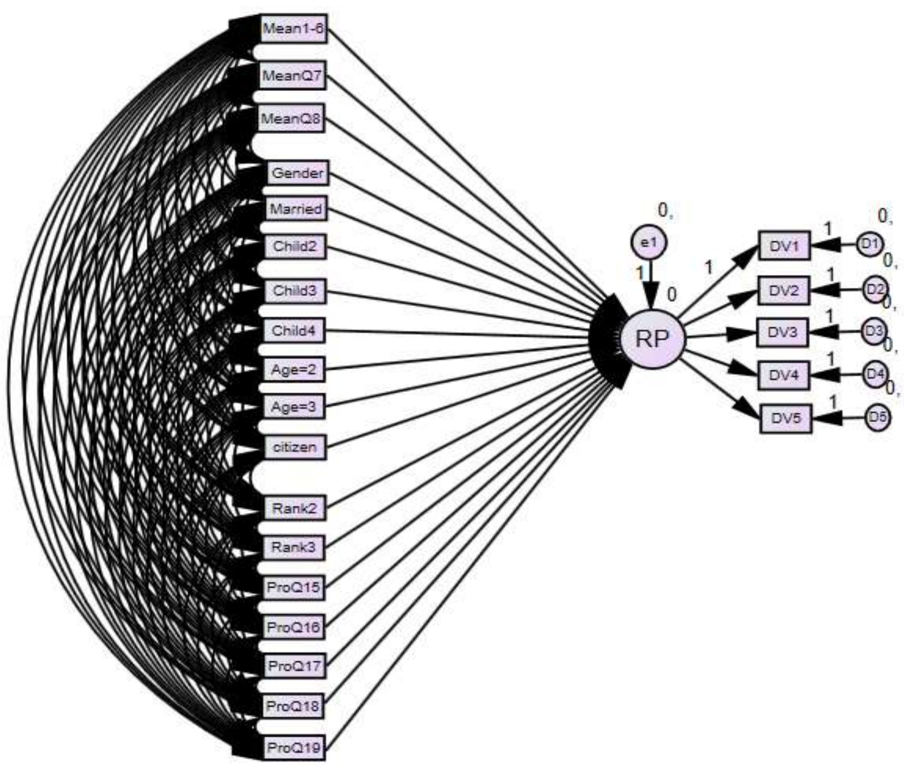

Figure 2. Model 2 with compressed data

Table 20 Results of SEM models (R^2) for original (Model 1) and compressed data (Model 2) sets

Dependent variables	Model 1	Model 2
Publication in journals	.004	.018
Published books	.728	.742
Published book chapters	.731	.743
Presenting papers	.126	.133
Edited and translated books	.906	.883
Total research productivity	.516	.364

Thus, both models 1 and 2 produced the same results in terms of nonsignificant β values but not for R^2. Previous analyses showed that model 2 yielded the best-fitting model among analyses performed with both the original and compressed data sets (CFI = .939, AIC = 599.020). Therefore, compressed data performed better than original data with SEM analysis but not with MR. Table 20 depicts the major findings of models 1 and 2 conducted with original and compressed data sets. Furthermore, comparing the results of model 2 to those analyzed with MR showed that SEM analyses produced dissimilar R^2 values and nonsignificant β values ($p > .05$). Table 21 displays the differences between SEM and MR analyses regarding R^2.

Table 21 Results of MR and SEM for compressed data, (R^2)

Dependent variables	Standard MR	SEM
Publication in journals	.266	.018
Published books	.275	.742
Published book chapters	.295	.743
Presenting papers	.160	.133
Edited books	.358	.883
Total RP	.305	.364

Overall, analyzing various models using MR and SEM with both original and compressed data sets yielded different results. Hierarchical, statistical, and standard models produced different results when analyzed by MR. Similarly, SEM analyses following the technique of using all causal indicators to predict the five indicators of faculty research productivity yielded different results. Thus, the examination of variability in the dependent measures was disparate among and between MR and SEM analyses.

Discussion

MR and SEM analyses were performed to examine whether they produced different answers to the question of whether personal, professional, and institutional variables predict the variation in faculty research productivity at Saudi universities. The results of this study revealed that the effect size, R^2, and significant individual predictors, β, varied depending on the analysis method used.

Hierarchical, standard, and statistical MR procedures were performed, which led to differential findings in terms of both the proportion of variability in dependent measures explained (R^2) as well as which individual predictors significantly affected the dependent variables (β).

Hierarchical MR was used to show the performance of each set of predictor variables, the change in R^2, and the overall significant R^2 when using different orders of model entry. According to theory, hierarchical regression is a suitable method of analysis because it shows the interactions among variables (Meyers et al., 2017). Indeed, several variables emerged as significant predictors in the final model but not in previous models and vice versa, which indicates that the possible presence of suppressor variables is correlated with other predictor variables but not criterion variables (Ludlow & Klein, 2014; Pedhazur, 1997). As a result, varying the order of entering variables into the models produced different significant individual predictors for each set of variables. Thus, although the overall models across the three procedures provided the same R^2 and significant β values as expected, the previous blocks that showed different results containing other significant individual predictors should be considered when testing models. This confirms Warner's (2013) conclusion that the strength and direction of the association among independent and dependent variables change based on the method of controlling other variables.

While statistical regression aims to include the variables that most strongly predict the criterion variables in the final model after beginning the analysis with no predictor variables, standard regression tests the performance of all predictor variables at the same time (Meyers et al., 2017). These regression methods vary because they use different techniques to process the independent variables in the regression equation and thus provide different explanations of the findings (Tabachnick et al., 2007). The R^2 values resulting from using statistical regression analysis on the dependent variables of publication in journals and published books were .165 and .351, respectively, both of which were smaller than those that resulted from using standard and hierarchical regression, which were .479 and .514, respectively. Further, some of the individual predictors that contributed to variation in the dependent variables in statistical regression were different than those in standard regression. Thus, the results showed that making a decision solely considering R^2 change may lead to inappropriate conclusions. Selecting different regression methods resulted in different outcomes, which should be considered by researchers, although these methods share a common goal in accounting for as much variance as possible in the outcome variable (Meyers et al., 2017).

Importantly, using SEM, which employs maximum likelihood estimation, overrides the problem of missing data shown in the MR analyses. Further, SEM has been chosen as an additional analysis because of the inability of MR to analyze models that include multiple dependent variables (Musil et al., 1998). Thus, the researcher used the MIMIC model to include both independent and dependent variables in a single model.

Kline (2006) and Petter et al. (2007) stated that researchers should consider their models to be formative measurement models when the assumption is that the latent variable in the model is caused by the predictor variables. However, an identification issue exists because of the impossibility of estimating the formative factor, the latent variable, when not including its outcomes, or reflective indicators

(Diamantopoulos, 2011). Accordingly, the researcher started analyzing the model that included 32 items predicting the latent variable (i.e., research productivity) that was measured by five dependent variables, or reflective indicators.

The first MIMIC model was analyzed by allowing all 32 formative indicators to covary in predicting the latent variable as measured by the five reflective indicators. This model was used based on the researcher’s hypothesis that these variables are somehow correlated, thereby estimating the covariance among all of them. This approach was supported by Jarvis et al. (2003) and Diamantopoulos (2011). Further, the latent variable, research productivity, was associated by the residual variance (i.e., disturbance), which takes the measurement error of variables into account; otherwise, this MIMIC model would be a linear combination of indicators predicting the Y-hat value in MR (Kline, 2006). The indices of goodness-of-fit for this model were $\chi^2(175, 133) = 279.654$, CFI = .914, RMSEA = .080, AIC = 1,493.654. Consequently, it is difficult to judge whether this model indicates a good or poor fit because the value of CFI falls between the values of .90 and .95 (Lai & Green, 2016). Although Franke et al. (2008) claimed that including reflective indicators (i.e., dependent variables) as functions of the latent variable is essential for the model to produce noteworthy estimations, the researcher found that excluding these dependent variables from the model did not affect the identification; rather, excluding them yielded a well-fitting model.

Additionally, with SEM analysis, none of the predictor variables produced significant β values, which shows disagreement with MR analyses. Furthermore, SEM produced R^2 values that contrast with those resulting from MR. It is impossible to conclude which analysis produced larger values of R^2 because their disparity did not follow a pattern. Further, Kline (2006) asserted the importance of specifying the measurement model by saying that assuming that variables are reflections of latent variables is often not a suitable approach. The results of this study indicate that the failure to distinguish between different measurement models can lead to misleading conclusions, a finding that confirms the results of the study conducted by Musil et al. (1998). Thus, these varying SEM models delivered different results, both among each other and compared with the MR analyses.

Limitations

The potential limitation that may affect the results in this study is the sample size. The researcher faced a challenge in data collection because universities do not allow researchers to access the email list of their faculty members; rather, surveys must be sent internally on behalf of the researcher. This challenge of reaching the sampling frame may be understood when considering Meyers et al.’s (2017) declaration that universities are one of the contexts for which researchers find a hard-to-reach population, thus making it hard to obtain a desirable sample size. Moreover, using a listwise deletion method with the existing survey, which was not under the control of the researcher, led to the small sample size. However, the researcher used dummy variable coding as a strategy to overcome the sample size issue, resulting in new data that included all cases.

Finally, although SEM theoretically requires large sample sizes, analysis with small sample sizes is used in the majority of SEM research (Kline, 2016). Further, Barrett (2007) emphasized that population properties, such as accessibility of the population, are key considerations when justifying the use of SEM in studies with sample sizes less than 200.

Conclusions

The researcher demonstrated whether the analytic choice between MR and SEM affected the results when examining the factors affecting research productivity of faculty members in Saudi Arabia. Though the results showed that choosing MR or SEM affected the results, it is worthy to consider the disparity in the results when selecting different methods or models among each analysis. A lack of understanding the differences between MR methods (i.e., statistical, standard, and hierarchical) affects the accuracy of conclusions because variables behave differently based on the method used. Therefore, the selection of the appropriate regression analysis should primarily depend on the nature of the research questions as well as whether the researchers are more interested in model testing or model building.

Importantly, the dissimilar results of various SEM models tested in this study affirm how specifying formative (i.e., casual) indicators incorrectly by treating them as reflective indicators results in worse data-fitting models, indicating the effect of model misspecification on the model fitting (Diamantopoulos et al., 2008). Thus, researchers should consider the distinction between formative and reflective measurement models when evaluating SEM models. Consequently, in terms of specifying models properly, researchers need to consider the effectiveness of their justifications, such as the latent variable corresponds to the function of the survey’s items.

Hence, the disparity in the results among various models and between MR and SEM methods may underscore the importance of very carefully selecting the appropriate analytic procedure. Researchers should ensure that the properties of each selected statistical procedure match both the goals of the research questions and the properties of the data, rather than simply choosing the more familiar method. Thus, the researcher found that it is critical for analysts to consider the advantage of each method and its ability to either affect the results or yield additional information that will contribute to the interpretation of the results.

Moreover, when MR equations are needed to investigate a single model, seemingly unrelated regression (SUR) may be a suitable method when accounting for correlated errors (Tan, 2018). Thus, future research may consider the application of SUR compared to MR or SEM for further investigation of both the importance of selecting the appropriate analytic approach and avoiding Type VI errors. Further, using SEM demands a large sample size, which is sometimes unrealistic. As such, another method that might be an alternative to SEM is factor score path analysis (i.e., the bias correcting method) (Devlieger & Rosseel, 2017). However, further investigation is needed to explore the contexts in which each method would work appropriately.

Acknowledgment:
I am greatly thankful for my advisor and dissertation chair, Prof. Gordon Brooks. I appreciate all his guidance, support, discussion, conversation, and advice regarding every aspect of my life that he has given me all along. I would like to give the most credits to Prof. Brooks for rigorously training me to become a statistician and a computer programmer.

References

1. Alzuman, A. (2015). Faculty Research Productivity in Saudi Arabian Public Universities:A Human Capital Investment Perspective (Doctoral dissertation). Retrieved from https://scholarscompass.vcu.edu/cgi/viewcontent.cgi?article=5065&context=etd
2. Barrett, P. (2007). Structural equation modelling: Adjudging model fit. Personality and Individual differences, 42(5), 815-824.
3. Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple

structure. Structural Equation Modeling, 12(1), 41-75.
4. Brown, T. A. (2015). Confirmatory factor analysis for applied research. Guilford Press.
5. Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. Sage Focus Editions, 154, 136-136.
6. Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. Journal of Business Research, 61(12), 1250-1262.
7. Devlieger, I., & Rosseel, Y. (2017). Factor score path analysis: An alternative for SEM? Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 13(Suppl 1), 31-38. http://dx.doi.org/10.1027/1614-2241/a000130
8. Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. Journal of business research, 61(12), 1203-1218.
9. Franke, G. R., Preacher, K. J., & Rigdon, E. E. (2008). Proportional structural effects of formative indicators. Journal of Business Research, 61(12), 1229-1237.
10. Grace, J. B., & Bollen, K. A. (2005). Interpreting the results from multiple regression and structural equation models. The Bulletin of the Ecological Society of America, 86(4), 283 -295.
11. Howell, R. D. (2014). What is the latent variable in causal indicator models? Measurement:Interdisciplinary Research & Perspectives, 12(4), 141-145.
12. Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. Journal of Consumer Research, 30(2), 199-218.
13. Kline, R. B. (2006). Reverse arrow dynamics: Feedback loops and formative measurement. In G. R. Hancock and R. O. Mueller, (Eds.), Structural Equation Modeling: A second course (pp. 43-68). Greenwich, CT: Information Age Publishing.
14. Kline, R. B. (2016). Principles and practice of structural equation modeling. Guilford Press.
15. Li, S. D. (2011). Testing mediation using multiple regression and structural equation modeling analyses in secondary data. Evaluation review, 35(3), 240-268.
16. Ludlow, L., & Klein, K. (2014). Suppressor variables: the difference between 'is' versus 'acting as'. Journal of Statistics Education, 22(2).
17. MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling inpsychological research. Annual review of psychology, 51(1), 201-226.
18. Meade, A. W. (2008). Power of AFI's to Detect CFA Model Misfit. In 23th Annual Conference of the Society for Industrial and Organizational Psychology, San Francisco, CA. Open acces via: http://www4. ncsu. edu/~ awmeade/Links/Papers/AFI (SIOP08). pdf.
19. Meyers, L. S., Gamst, G., & Guarino, A. J. (2017). Applied multivariate research: Design and interpretation. Los Angeles: SAGE.
20. Musil, C. M., Jones, S. L., & Warner, C. D. (1998). Structural equation modeling and its relationship to multiple regression and factor analysis. Research in Nursing & Health, 21(3), 271-281.
21. Nachtigall, C., Kroehne, U., Funke, F., & Steyer, R. (2003). Pros and cons of structural

equation modeling. Methods Psychological Research Online, 8(2), 1-22.
22. Nayak, B. K., & Hazra, A. (2011). How to choose the right statistical test?. Indian Journal ofOphthalmology, 59(2), 85.
23. Pedhazur, E. J. (1997). Multiple regression in behavioral research: Explanation and prediction. Belmont, Ca.: Wadsworth.
24. Petter, S., Straub, D., & Rai, A. (2007). Specifying Formative Constructs in Information Systems Research. MIS Quarterly, 31(4), 623–656. https://doi org.proxy.library.ohio.edu/10.2307/251488 14
25. Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of Fit Measures. Methods Of Psychological Research, 8(2), 23-74.
26. Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., ..., & Carlsson, R. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. Advances in Methods and Practices in Psychological Science, 1(3), 337-356.
27. Simundic, A. M. (2016). Statistical analysis in method comparison studies – Part one.
28. Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). Using multivariate statistics (Vol. 5). Boston, MA: Pearson.
29. Tan, L., Chiong, K. X., & Moon, H. R. (2018). Estimation of High-Dimensional Seemingly Unrelated Regression Models. arXiv preprint arXiv:1811.05567.
30. Wang, J., & Wang, X. (2012). Structural equation modeling: Applications using Mplus. John Wiley & Sons.
31. Wang, X., French, B. F., & Clay, P. F. (2015). Convergent and discriminant validity with formative measurement: A mediator perspective. Journal of Modern Applied Statistical Methods, 14(1), 11.
32. Warner, R. M. (2013). Applied statistics: From bivariate through multivariate techniques. Sage Publications.
33. Westland, J. C. (2010). Lower bounds on sample size in structural equation modeling. Electronic Commerce Research and Applications, 9(6), 476-487.