Studies in Systems, Decision and Control 285

Jitesh J. Thakkar

Structural Equation Modelling

Application for Research and Practice (with AMOS and R)

Springer

Studies in Systems, Decision and Control

Volume 285

The series "Studies in Systems, Decision and Control" (SSDC) covers both new developments and advances, as well as the state of the art, in the various areas of broadly perceived systems, decision making and control–quickly, up to date and with a high quality. The intent is to cover the theory, applications, and perspectives on the state of the art and future developments relevant to systems, decision making, control, complex processes and related areas, as embedded in the fields of engineering, computer science, physics, economics, social and life sciences, as well as the paradigms and methodologies behind them. The series contains monographs, textbooks, lecture notes and edited volumes in systems, decision making and control spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

** Indexing: The books of this series are submitted to ISI, SCOPUS, DBLP, Ulrichs, MathSciNet, Current Mathematical Publications, Mathematical Reviews, Zentralblatt Math: MetaPress and Springerlink.

More information about this series at http://www.springer.com/series/13304

Jitesh J. Thakkar

Structural Equation Modelling

Application for Research and Practice (with AMOS and R)

Springer

Jitesh J. Thakkar
Department of Industrial and Systems
Engineering
Indian Institute of Technology Kharagpur
Kharagpur, West Bengal, India

ISSN 2198-4182 ISSN 2198-4190 (electronic)
Studies in Systems, Decision and Control
ISBN 978-981-15-3792-9 ISBN 978-981-15-3793-6 (eBook)
https://doi.org/10.1007/978-981-15-3793-6

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Dedicated to
my guru
Prof. S. G. Deshmukh

Preface

Structural Equation Modelling (SEM) is a multivariate quantitative technique employed to describe the relationships among observed variables. The technique helps the researcher to test or validate a theoretical model for theory testing and extension. Typically, the interest of a researcher is to investigate the constructs emerging out of sets of variable and how these constructs are related to each other. For example, a sales manager might be interested to investigate a phenomenon that behaviour and discipline of salesperson have a direct influence on sales volume. Similarly, a researcher may hypothesize that the overall fitness of the player influences his/her performance in the sports. An oncologist might be interested to verify that a too much intact of protein leads to breast cancer in female. It is the preliminary understanding of the researcher that in some way the sets of variables that define the constructs are related in a certain way. As a result, the researcher wants to verify that to what extent the hypothesized theoretical model is adequate for the sample data. On verification of this, the researcher gets exposed to two options: (i) if the hypothesized theoretical model is supported by the sample data, then a researcher can incorporate more phenomena in the basic model and attempt to investigate more complex structure; (ii) if the theoretical model is not adequately supported by the data, then a researcher should either modify the basic model or develop an alternative model for testing. SEM enables the researcher to indulge into a deeper inquiry through a process of scientific hypothesis testing and extend the present body of knowledge by discovering complex relationships among constructs. This book provides a comprehensive learning on SEM for the academic and industry researchers. The key features of the book are:

- A comprehensive book for researchers and industry professionals
- Conceptual and mathematical understanding of SEM
- Illustrative step-by-step applications of SEM with AMOS, SPSS and R software programs

- Summary of research applications of SEM in operations management, psychology, humanities, human resources, organizational behaviour, marketing
- Glossary
- Important video links on SEM
- Frequently asked questions on use of SEM.

Kharagpur, India
February 2020

Dr. Jitesh J. Thakkar

Acknowledgements

It gives me immense pleasure to deliver this book on *Structural Equation Modelling* to students. This book will benefit students/researchers in engineering, management and social science field. The field of structural equation modelling has received the contributions from many scholars; hence, these individuals and scholars have contributed to the development of this book.

The discussion surrounding this topic and growing research applications of SEM has helped shape this book. I have always received inspiration and energy for executing academic projects from my teacher and guru Prof. S. G. Deshmukh. I express my deep gratitude for his valuable direction and inputs without which this book would not have attained its present form, in both content and presentation. He has been a supportive mentor, and it is with his help that I have achieved both professional and personal successes.

I acknowledge the support extended by the Department of Industrial and Systems Engineering, Indian Institute of Technology Kharagpur, in executing this book-writing project. A work culture and passion of my faculty colleagues for academic excellence have always inspired me. I am thankful for the support of my students in coding and verification of the models.

My father, Jayprakashbhai Thakkar, and mother, Ushaben Thakkar, have provided a constant moral support and motivation for this work. I deeply express my love and affection for my wife, Amee, daughter, Prachi, and son, Harshit, for giving me freedom and moral support for the completion of this book.

Kharagpur, India
February 2020

Dr. Jitesh J. Thakkar

Contents

About the Author

Dr. Jitesh J. Thakkar is a faculty in the department of Industrial and Systems Engineering at IIT Kharagpur. He obtained his Ph.D. and M.Tech. from IIT Delhi where he worked in the area of Industrial Engineering. He received Bachelors in Mechanical Engineering with Gold Medal from one of the oldest Engineering College Birla Vishvakarma Mahavidyalaya, Sardar Patel University, Gujarat. He has 20 years of teaching, research and industry experience. He has guided six PhDs in the areas of lean manufacturing, sustainable supply chain management, and service operations management at IIT Kharagpur. He has published more than 60 research papers in SCI and SCOPUS indexed journals in the areas of Industrial Engineering and Operations Management. He is a productive researcher with H-index - 22 and total citations more than 2000. His publications have appeared in the high impact factor journals such as International Journal of Production Economics, Transportation Research (Part-E), International Journal of Production Research, Computers and Industrial Engineering, Production Planning and Control, Expert Systems with Applications, Journal of Cleaner Production, International Journal of Advanced Manufacturing Technology, Electronic Commerce Research and Applications, Resources Policy, Journal of Manufacturing Technology Management and International Journal of Productivity and Performance Management. He is rendering his services as an editorial board member of the journals - OPSEARCH, Journal of Manufacturing Technology Management, International Journal of Quality and Reliability Management, and International Journal of Productivity and Performance Management. He has trained Corporate Managers in Lean Manufacturing, Process Excellence, Six Sigma, Project Management, Quality Management, Supply Chain Management and Statistical Decision Making. His online (SWAYAM-NPTEL) course on “Six Sigma” has benefited more than 20,000 University students and industry professionals. He has been invited as a faculty expert by various reputed institutes such as IIT Kanpur, IIT Madras, IIM Indore, NITIE Mumbai, NIT Surat, NIT Trichy, Institute of Rural Management Anand, Ahmedabad Management Association, BCCI Kolkata, and L&T Project Management Institute, Chennai.

Chapter 1
Introduction to Structural Equation Modelling

1.1 What Is Structural Equation Modelling (SEM)?

Structural equation modelling (SEM) is multivariate quantitative technique employed to describe the relationships among observed variables. The technique helps the researcher to test or validate a theoretical model for theory testing and extension. The multivariate analysis is conducted with an objective to help the researcher for an in-depth explanatory analysis with a required statistical efficiency. A researcher is interested in investigating the constructs emerging out of sets of variables and the relationships among these constructs. This can be explained with an example such as a sales manager might be interested to investigate a phenomenon that behaviour and discipline of salesperson has a direct influence on sales volume. Similarly, a researcher may hypothesize that the overall fitness of the player influences his/her performance in sports. An oncologist might be interested to verify that a too much intact of protein leads to breast cancer in female. A researcher tries to investigate the relationships among the set of variables defining a particular construct. As a result, the researcher wants to verify that to what extent the hypothesized theoretical model is adequate for the sample data. On verification of this, the researcher gets exposed to two options: (i) if the hypothesized theoretical model is supported by the sample data, then a researcher can incorporate more phenomena in the basic model and attempt to investigate more complex structure; (ii) if the theoretical model is not adequately supported by the data, then a researcher should either modify the basic model or develop an alternative model for testing. SEM enables the researcher to indulge into a deeper enquiry through a process of scientific hypothesis testing and extend the present body of knowledge by discovering complex relationships among constructs. It is difficult to test the complete theory of a researcher which can consider an extensive causality among the constructs with other multivariate techniques. In this regard, SEM should be seen as an extension to other multivariate techniques, specifically such as factor analysis and multiple regression analysis. SEM is considered to be a very much useful technique for examining series of dependence relationships simultaneously by solving multiple equations. For example, we may

J. J. Thakkar, *Structural Equation Modelling*, Studies in Systems, Decision and Control 285, https://doi.org/10.1007/978-981-15-3793-6_1

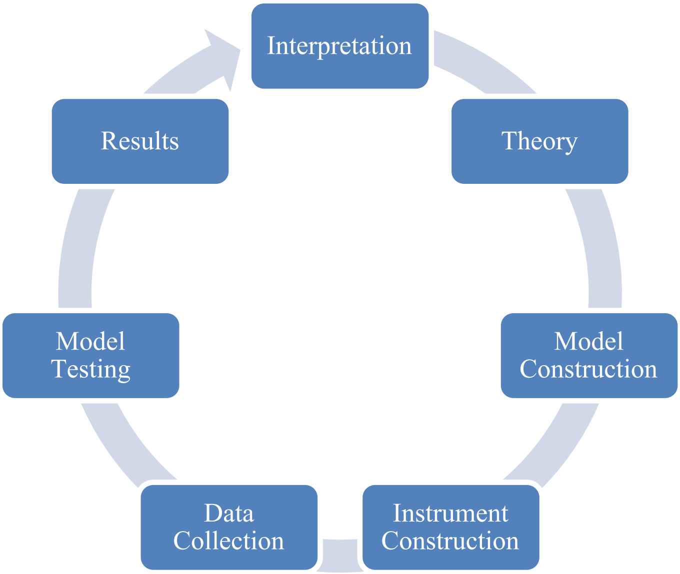


Fig. 1.1 General approach for conducting a SEM analysis

assume that patient recovery leads to better Return on Investment (ROI) for a hospital system and better ROI helps to improve infrastructure and hence enhances patient recovery. In this relationship, we can see that ROI is acting as both dependent and independent variables. It is not possible to accommodate the dual nature of one variable and its causal relationships with other variables in one multivariate model only. An overview of SEM process is presented in Fig. 1.1.

There are select key issues in SEM modelling. The preliminary understanding of this will enable the reader to follow the model building, testing and interpreting process in a better way. Typically, this includes the following concepts:

- **Latent variable (constructs or factors)**: These variables are not directly observed or measured. The inference of latent variable is indirect, and typically, it is based on the set of factors that are measured using surveys, tests, etc. For example, faith in a particular spiritual sector is a latent variable. The confidence of the student is a latent variable that represents behaviour and academic performance of the student. An individual intelligence can be considered as latent variable. The latent variables are related to measured variables with a dependence relationship.
- **Observed variable (constructs or factors)**: These variables are used to define or measure or interpret the latent variables. For example, blood pressure and sugar level are the observed constructs that represent the overall health of the person. The Semester Grade Performance Index (SGPI) is a measure of students' overall academic discipline and interest in the studies.
- **Independent or dependent variables**: The observed or latent variable can be independent (not influenced by other variables) or dependent nature (influenced by

another variable). For example, productivity is an independent construct or variable which is influenced by many other dependent variables such as: technology, HR policies, employee motivation.

- **Regression model**: It predicts single dependent observed variable based on one or more independent observed variables. For example, set of independent variables, such as education, experience, researcher publications, of teachers in a typical university can be used to predict the overall score and rank of the university.
- **Path model**: It intends to test more complex models compared to regression model by considering multiple independent and dependent observed variables. For example, students' academic performance, attendance in the class, job placement, etc. can be used to quantify the performance of university, student satisfaction level and brand value of the institute.
- **Confirmatory factor models**: This model includes observed variables. These variables intend to measure one or more latent variables (independent or dependent). For example, education and experience of doctors, infrastructural capacity and expertise in various areas are observed variables of the independent latent variable "brand image of hospital".
- **Structural equation models**: It includes both observed and latent variables, which may be of independent or dependent types. For example, an independent latent variable (self-discipline) influences a dependent latent variable (academic performance). Here, both types of latent variables (independent and dependent) are measured, defined or inferred by multiple observed or measured indicator variables. It is to be noted that path analysis is considered as a special case of SEM. Path analysis includes only observed variables and has a more stringent set of assumptions compared to SEM. The fundamental difference between SEM and path analysis lies in the assumption about error. Path analysis assumes the measurement of all the variables without error, whereas SEM uses latent variables to account for measurement error. Typically, error in SEM is considered of two types. First type of error is systematic error. It is primarily because of bias in the responses collected through questionnaire. For example, a questionnaire collecting data on the patient satisfaction in a multispeciality hospital may get responses which are rated on the higher or lower side because of nature of the questions or manner in which the questionnaire is administered. In case of a systematic error, the mean of errors does not get cancelled out (mean of error $\neq 0$). Second type of error is random error. In this case, mean of the errors gets cancelled out (mean of error $= 0$). The observed value of the variable in SEM is influenced by true score of the variable and error term. It is not possible to solve the equation and identify the value of observed variable when we have two unknown parameters like true score and error. For this purpose, it is necessary to include multiple indicators of the latent variable in SEM analysis. This can help to estimate the values of true score and error.

Structural equation modelling (SEM) is a powerful multivariate statistical tool which is used to test and evaluate the causal relationship in a model. It evaluates the hypotheses to relate observed and unobserved (latent) variables. It is a used to represent, estimate and test the theoretical linear framework among variables. It is used to test hypothesized forms of directional and non-directional relationship between measured variables and latent variables. Structural modelling technique is about 100-year-old technique and which has been developed over three generations. The logic of casual modelling was developed in the first generation. In the second-generation, factor analysis was included by the social sciences and hence expanded its capacity. The third generation of SEM started in 2000 with Judea Pearl's development of the "structural causal model", which was followed by Lee's in 2007 which was basically the integration of Bayesian modelling in SEM. The two aims in conducting a SEM analysis are:

(1) It helps to understand the correlation or covariance pattern between a variable set.
(2) It gives an explanation about their variance as possible with the specified model.

1.2 A Comparison of Traditional Statistical Techniques and SEM

Structural equation modelling is quite similar to the traditional techniques like correlation, variance analysis and regression. Firstly, both SEM and traditional methods have base of linear statistical model. Secondly, for conducting statistical analysis, both methods have some basic assumptions to be followed as traditional methods have the assumption of normal distribution and SEM has assumption of multivariate normality. Lastly, both approaches do not offer test of causalities. There are few differences between traditional method and SEM method as explained in Table 1.1.

1.3 Brief History of Causality Models

The key milestones in the history of causality are outlined as below:

- Analysis of variance (1920–1930): It decomposes the variance of a dependent variable. This helps to identify the contribution of an explanatory variable.
- Macro econometric models (1940–50): It includes all the variables in the model for a dependence analysis of non-experimental data.
- Path analysis (1920–70): It is similar to econometric models but primarily focuses on the analysis of correlations.
- Factor analysis (1900–1970): It analyzes the correlations among multiple indicators of the same variable.
- SEM (1970): This has integrated econometric models, path analysis and factor analysis. It employs an interdependence analysis to evaluate the relationships among variables.

Table 1.1 Differences between traditional statistical techniques and SEM

S. No.	Basis	Traditional statistical techniques	Structural Equation Modelling (SEM)
1	Flexibility	Not much flexible	Highly flexible and comprehensive methodology
2	Nature of model	Default model is specified	No default model is offered. It puts selects restrictions on the types of relations that can be specified. SEM model specifies relations a priori by considering formal specifications of a model. This helps researchers to support hypothesis with theory or research
3	Variable analysis	Analyze only measured variables	It is a multivariate technique which incorporates the variables such as observed (measured) and unobserved (latent constructs)
4	Errors	Assume measurement occurs without error	SEM explicitly specifies error to provide researcher a flexibility in considering imperfect nature of their measures
5	Significance test	It determines group differences, relationships between variables, or the amount of variance explained using significance tests	It does not provide any straightforward tests to determine model fit. The various tests are used to test the model fit. This includes comparative fit index (CFI), chi-square, Bentler-Bonett non-normed fit index (NNFI), root mean square error of approximation (RMSEA)
6	Approach	It does not have a graphical approach	The complex relationships in SEM are presented in a graphical language which provides an intuitive appeal and convenience in understanding the model. The model is specified by formulating statements about a set of variables. To test model fit and estimate parameters, the graphical representation of the model is transformed into a set of equations which are solved simultaneously

The chronological development of various models developed in the past 100 years is explained in Table 1.2.

1.4 Types of Structural Equation Models

There are many powerful SEMs which are unexplored yet and are highly flexible. Some of the variants of SEMs are explained below.

Table 1.2 History of causality models

Timeline	Model	Key contributors	Key features
1896	Linear regression models	Karl Pearson (correlation coefficient)	• The regression weights are calculated based on correlation coefficient and least squares criterion • It enables the prediction of dependent observed variable scores (Y) (e.g. sales of mobiles or no. of patients to be admitted, etc.)
1904–1927	Factor analysis	Charles Spearman	• Determines which items correlated to create the factor model • It helps to define and measure a construct based on correlated items
1940	Factor techniques	D. N. Lawley and L. L. Thurstone	Instruments (sets of items) that yield observed scores from which inference about the constructs is made
1955–1965	confirmatory factor analysis (CFA)	Howe (1955), Anderson and Rubin (1956), and Lawley (1958) Karl Jöreskog (1960) Jöreskog (1963)	• Tests whether a set of items defined a construct • Used to create measurement instruments used in many academic disciplines • Used to test the existence of the theoretical constructs

(continued)

Table 1.2 (continued)

Timeline	Model	Key contributors	Key features
1918–1960	Path models	Sewell Wright (1918–1934) H. World (1950s) D. Duncan and H. M. Blalock (1960s)	• The complex relationships among observed variables are investigated using correlation coefficients and regression analysis • The relationship among the observed variables in the path model is established by solving a set of simultaneous regression equations
1973–1994	Structural equation modelling (SEM)	Karl Jöreskog (1973) Ward Keesling (1972), and David Wiley (1973) Jöreskog and van Thillo (LISREL in 1973)	• It combines both path models and confirmatory factor models to incorporate both latent and observed variables • Initially known as the JKW model, but the development of LISREL software in 1973 gave it a unique identity like linear structural relations model (LISREL)

1.4.1 Latent Growth Curve (LGC) Model

Latent growth curve models are used to evaluate sample which has a serial change with time. It is constructed on the assumption that along with the sample series there is a structure growing along. User specifies the loading factors which are the series of growing subjects that are represented by the slope of growth which is a latent variable. LGC uses the longitudinal data with more than three periods rather than a time series analysis, which requires a larger time series/more observation. It is based on the assumption of a stable growth curve of the observation. This permits the users to weigh the curve based on the time span rather than time series. This entails steady intervals in the series.

1.4.2 Bayesian SEM (BSEM)

It is based on the assumption that the theoretical support and preceding beliefs are concrete. Researcher can use new sample to update a previous model so that posterior parameters can be evaluated. The basic benefit of using BSEM is that it does not need a data size or sample size but requires prior information on sample parameters and its distribution. For example, Arhonditsis et al. (2007) investigated spatiotemporal phytoplankton dynamics, with a sample size of <60 using BSEM. The parameters' posterior distribution is estimated based on various Monte Carlo simulations to compute the overall mean and a 95% confidence interval. The Bayesian framework enables the model assessment of BSEM more like a model comparison which is not typically based on χ^2, RMSEA, CFI, etc. The SEM analysis, which uses maximum likelihood (ML) and the likelihood ratio χ^2 test, unnecessarily utilizes model modification and often strictly rejects the substantive theory to improve the model fit by chance. This has increased the attention of researchers in the use of Bayesian approach in SEM applications due to its flexibility and better representation of the theory.

1.4.3 Partial Least Square SEM (PLS-SEM)

PLS-SEM is mostly recommended when a researcher does not have a well-developed theoretical base, especially when there is minimal earlier information on causal relationship. The importance of this technique is given to exploration than to confirmation as it does not require large data set or assumption on distribution of data. Therefore, people with small data sets and less theoretical background can use PLS-SEM to test for causality for their research (Hair et al. 2014). The algorithms for PLS-SEM have the basis of maximum likelihood. It is recommended that when researchers do not have a sufficient data, then first go for PLS-SEM to construct a proof for casual relationship among variables. This helps the researcher to remain collecting their long term sample while researcher can also update their hypothesis.

1.4.4 Hierarchical SEM

Hierarchical SEM is also known as multilevel SEM which analyzes clustered sample in a hierarchical manner. It can help to lay down direct and indirect causal relationships among cluster. The traditional SEM technique ignores the fact that the intercepts and the path coefficients will possibly vary between hierarchical levels. Hierarchical SEM concentrates on sample produced with a hierarchical form. Hence, the data size should be huge. This SEM method is very flexible.

1.5 SEM Software Programmes

The present market offers various options to a researcher for conducting SEM analysis in the well-designed and user-friendly software programmes. The present softwares primarily offer four features: (i) it has a capability of conducting statistical analysis of raw data (means, correlations, missing data conventions, etc.); (ii) provide routines for handling missing data and detecting outliers; (iii) generate the programme's syntax language, diagram the model, and (iv) provide an easy import and export of data and figures of the theoretical model(s). The recent software is equipped with the sets of sample data and provides detailed instructions in their manual with an illustrative example for conducting a step-by-step SEM analysis. The summary of the key SEM software packages is presented in Table 1.3.

1.6 Reasons for Popularity of SEM

1. The popularity of SEM is exponentially growing in various fields like social science, humanities, engineering, marketing, behavioural sciences and other fields. This is primarily because of its ability to simultaneously estimate the multiple equations by considering the relationships between constructs and measured

Table 1.3 Key features of SEM software

S. No.	Software	Key features
1	Amos (SPSS interface)	• It provides an easy to use graphical interface to visually construct models with common online drawing tools • Can build attitudinal and behavioural models that reflect complex relationships more accurately than with standard multivariate statistics techniques using either an intuitive graphical or programmatic user interface • It has easy-to-use interface for bootstrapping methods, which can be applied to parameter estimates, effect estimates, sample means, sample variances and covariances, correlations, model comparisons, and comparisons of estimation methods • It can accommodate non-recursive models, models with fixed parameters and models based on data from multiple populations
2	EQS	• SEM model can be created in different ways and extends greater capability for exploratory analysis • Handles non-normal variables in complete and missing data situations • In includes various features necessary for conducting SEM analysis. This includes multiple regression, multivariate regression, confirmatory factor analysis, structured means analysis, path analysis and multiple population comparisons

(continued)

Table 1.3 (continued)

S. No.	Software	Key features
3	LISREL-SIMPLIS, LISREL-PRELIS, Interactive LISREL	• Accommodates the modelling of linear and nonlinear hierarchical models and the evaluation of a model in different study groups (group comparison) • Easy to import the external data in various formats like SPSS, SAS, MS Excel, etc. • Handle models with measurement error • Help the researcher in non-recursive models • Useful for working on multigroup comparisons (like developing separate models for males and females, etc.) • It is used for decomposition of certain effects that are initially done manually by the researcher
4	Mplus	• Provides an integrated modelling framework to handle continuous, categorical, observed, and latent variables • It incorporates latest missing data handling methods • Incorporates auxiliary variables for missing data handling • Unique capability including Bayesian SEM, exploratory SEM, handling of count and censored values, and missing data under not missing at random (NMAR) mechanism
5	SAS PROC CALIS	• Flexibility of model specification in many different formats such as COSAN, FACTOR, LINEQS, MSTRUCT and RAM • Improved mean structures analysis • New and improved modelling languages • Multiple-group analysis • Improved standardized results • Improved effect analysis
6	R package sem	• It is an open-source environment with complete features of R including graphics capabilities • Incorporation of two-stage least squares for econometric modelling • Ability to create path diagram via Graphviz
7	R package lavaan	• It is an open-source environment with complete features of R including graphics capabilities • Provides compact scripting for complicated models • Ability to compute two different likelihoods for ML estimation (normal and Wishart) • Ability to "mimic" Mplus and EQS results
8	R package OpenMx	• It is an open-source environment with complete features of R including graphics capabilities • Ability to optimize user-specified objective functions Hierarchical model definition to link a series of models in a tree structure Ability to use the programme as an optimizer and matrix algebra calculator

indicator items. It also considers the associations among constructs under investigation. This is performed by conducting factor analysis and regression analysis in one step only.

2. Today, the dynamic nature of enquiry forces the researcher to accommodate causality among various latent and observed variables of both independent and dependent type. This helps them to gain greater insights into their area of research. The limited number of variables can be analyzed with the help of basic statistical methods and does not enable the development of sophisticated theories. For instance, it is not adequate to use simple bivariate correlations for examining a sophisticated theoretical model. In contrast, structural equation modelling permits the complex phenomena to be statistically modelled and tested. This makes SEM preferred and most demanding approach for confirming (or discontinuing) theoretical models with an accurate quantification of the variables. SEM helps researchers and analysts to simultaneously examine the series of interrelated dependence relationships among the multiple variables such as measured variables and latent constructs (variates). It also examines the relationships among several latent constructs. SEM provides a unique integration interdependence and dependence techniques by combining the use of two multivariate techniques such as factor analysis and multiple regression analysis.
3. It has been observed that a greater recognition is given to the validity and the reliability of observed scores from measurement instruments. Specifically, measurement error has become a major issue in many disciplines, but measurement error and statistical analysis of data have been treated separately. Structural equation modelling techniques explicitly take measurement error into account when statistically analyzing data. As noted in subsequent chapters, SEM analysis includes latent and observed variables as well as measurement error terms in certain SEM models.
4. SEM has been evolved over a period of 40 years. This has helped the researchers to incorporate many advanced features in SEM analysis. For example, group differences in theoretical models can be assessed through multiple-group SEM models. In addition, collecting educational data at more than one level, for example, from students, teachers and schools, is now possible using multilevel SEM modelling. As a final example, interaction terms can now be included in an SEM model so that main effects and interaction effects can be tested. These advanced SEM models and techniques have provided many researchers with an increased capability to analyze sophisticated theoretical models of complex phenomena, thus requiring less reliance on basic statistical methods.
5. The user-friendliness of SEM software programmes is one of the reasons for its popularity in the contemporary researchers. SEM software programmes are Windows based and use pull-down menus or drawing programmes to generate the programme syntax internally. This makes the use and adoptability of SEM software easier.

Chapter 2
Technical Aspects of SEM

2.1 Basic Terminology of SEM

It is important to get acquainted with the key terminologies used in SEM. This mainly includes

1. **Exogenous Constructs/Variables**: They are those variables which are not influenced by any other variables present in the model. For instance, there are two factors that cause changes in percentage of student such as number of hours studied per month and IQ, and they both do not have any causal relationship among them. Then, IQ and hours studied per month will be exogenous variable. An exogenous construct has only correlational relationships with other constructs (i.e. no dependence path coming into construct).
2. **Endogenous Constructs/Variables**: They are those variables which are influenced by any other variables present in the model, such as percentage of the student in previous example will be endogenous variable. It is a dependent variable. In SEM model, any construct with a dependence path (arrow) pointing to it is considered endogenous.
3. **Manifest Variable**: A variable which is observed and measured directly is known as manifest variable. Manifest variable is also known as the indicator variable. In the above example, all the variable present can be directly monitored, and hence, they are called as manifest variable. In SEM when we only examine manifest variable, then the model is called Path Analysis.
4. **Latent Variable**: A variable which cannot be observed and measured directly is known as latent variable. Latent variable in factor analysis is known as factors. For instance, in the previous example, we also need to calculate the effect of motivation on overall percentage, and then, it is the latent variable. Motivation cannot be quantified, it is internal hence can only be quantified on the basis of response of the questionnaire by the student, and hence, it is a latent variable. It increases the complexity of SEM as one needs to consider all the questionnaire items and has to measure the responses that are used to quantify the latent variables or the "factors".

J. J. Thakkar, *Structural Equation Modelling*, Studies in Systems, Decision and Control 285, https://doi.org/10.1007/978-981-15-3793-6_2

5. **Moderation**: In a typical situation involving three or more variables, when the presence of one variable changes the dynamics of other two variables then the situation is known as moderation. Therefore, moderation exists only when relationship between two variables is not at the same level with the third variable. One way to think of moderation is when you observe an interaction between two variables in an ANOVA.
6. **Covariance** and **correlation** are the building blocks of how your data will be represented when doing any programming or model specification within a software program that implements structural equation modelling. You should know how to obtain a correlation matrix or covariance matrix using PROC CORR in SAS or use other menu tools from a statistical package of your choice, to specify that a correlation or covariance matrix be calculated. The covariance matrix in practice serves as your dataset to be analysed. In the context of SEM, covariances and correlations between variables are essential because they allow you to include a relationship between two variables that is not necessarily causal. In practice, most structural equation models contain both causal and non-causal relationships. Obtaining covariance estimates between variables allows one to better estimate direct and indirect effects with other variables, particularly in complex models with many parameters to be estimated.
7. **Measurement model**: It assigns indicator variables to the constructs they should represent. It is essential to develop measurement model when you need to incorporate latent variables in your model. It relates the measured variables to latent variables and examines the relationships between them. It is necessary to verify select issues before constructing a measurement model. Typically, this includes

   - What is the direction of influence between the latent variable (causal indicators) and indicators (effect indicators)?
   - What is the nature of indicators—causal or effect?
   - Do we have single or multiple indicators of the latent variable?
   - What is the nature of latent variable—non-continuous or continuous?
   - Is there a single or multiple latent variables which influence the indicator?
   - Do the measures intend to describe the construct or explain the construct (such as the set of constructs can be combined to represent an index)?

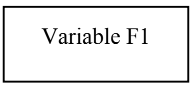
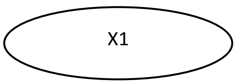
The measurement model in SEM analysis seems to be a replica of factor analysis. Typically, this represents loading on the factors to examine the strength of the relationship of each variable to the construct (known as loading in factor analysis). However, there is a key difference in the implementation of these two approaches—factor analysis and SEM. Factor analysis which typically examines the structure among variables and defines the factors in terms of set of variables is considered as exploratory analysis. In contrast to this, SEM demands an analyst to define the association of variables with each construct. Subsequently, loadings are estimated for the cases where variables are associated with constructs. A researcher needs not to develop any specification for exploratory analysis, whereas SEM demands complete specification of the measurement model.

8. **Structural model**: It relates all of the variables (both latent and manifest) that we need to account for in the model. The unique feature of SEM is a structural model which links the hypothesized constructs of the model to specify set of dependence relationships. This helps to estimate a series of separate but interdependent, multiple regression equations simultaneously. The key lies in analysing the set of relationships among the inter-connected variables simultaneously instead of examining them separately usually being done in other multivariate techniques. It is always recommended to first test the fit and construct validity of the proposed measurement model, and then as a second step, a researcher should test the structural model. The purpose of this model is to define the structural relationships among the constructs. It helps the researcher to investigate the underlying hypotheses by investigating the dependence relationships among the constructs.
9. **Recursive Model**: In a model when causation is only in one direction, then the model is known as recursive model.
10. **Non-recursive Model**: In a model when causation is both sides that means the flow is in both the directions such as giving a feedback, and then, the model is known as non-recursive model.

2.2 Basic Symbols and Relationships Used in Structural Equation Modelling

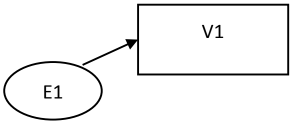
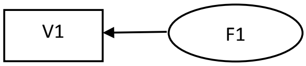
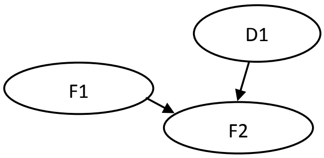
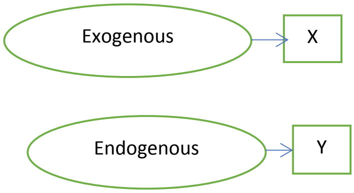
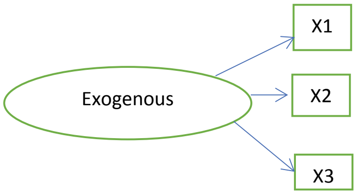
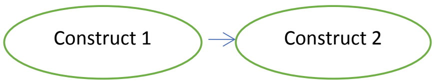
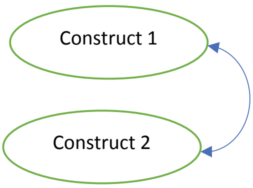
To make path modelling easy, structural equation modelling uses some basic symbols which are described in Table 2.1.

Table 2.1 Symbols used for conducting a SEM analysis

Symbols and relationships in SEM	Description
	Observed or measured variable
	Unobserved or unmeasured latent construct
→ ←	Direct relationship
↔	Covariance or correlation

(continued)

Table 2.1 (continued)

Symbols and relationships in SEM	Description
	Error $E1$ associated with measured variable $V1$
	Path coefficient on latent variable $F1$ on an observed variable $V1$
	Represents path coefficient for regression of one latent variable $F1$ onto another latent variable $F2$. Residual error in prediction of $F2$ by $F1$ is denoted as $D1$
	Indicates relationship between a construct and a measured variables
	It indicates relationship between a construct and multiple measured variables
	It indicates a structural relationship (dependence relationship between two constructs). Typically, measures impact of the impact of one construct on another construct or variable
	It represents the relationship between constructs

2.3 Path Analysis

It quantifies the relationship between several variables. SEM was also known as path analysis before the presence of latent variables. It was used to test the structural hypothesis of both direct and indirect causal relationship. Path analysis uses a common function known as mediation. Mediation is a technique in which one variable can influence an outcome indirectly or directly with the help of another variable. It is represented with nodes that imply variables and arrows showing relationship among these variables. In a path diagram, latent variable is represented by ellipse, measured variable is represented by rectangle or square, and relationship between arrows is represented by arrows. The direction of causal relationship is represented from a base of then arrow to head of the arrow. When two single-headed arrows facing each other opposite in direction are present, then it shows the feedback of causal relationship. Association among two variables is represented by curved two-headed arrow. An arrow is drawn from the value of the error term to the variable with which the term is associated to represent error terms for a variable.

2.4 Mathematical Explanation with Example

Once the model is specified, a researcher needs to choose an appropriate estimation method to estimate the relationships in SEM model. Initially, SEM estimations were carried out with ordinary least squares (OLS) regression. Subsequently, researchers have started using an alternate efficient and unbiased estimator under the assumption of multivariate normality which is maximum likelihood estimation (MLE). This estimator was widely in the initial version of LISREL program. However, this estimator is very much sensitive to non-normality, and hence, the use of this estimator under non-normality conditions is replaced by select estimation methods such as weighted least squares (WLS), generalized least square (GLS) and asymptotically distribution-free (ADF) estimation. Specifically, ADF has become widely popular for its very low sensitivity to non-normality of the data (Fig. 2.1).

Equations (2.3)–(2.7) indicate the directional influences of the exogenous LVs (ξ) on their indicators (x). Thus, Eq. (2.8) links the observed (manifest) variables to unobserved (latent) variables through a factor analytic model and constitutes the “measurement” portion of the model. Equation (2.9) is final measurement equation.

$$
\begin{array}{cccccc}
 & \multicolumn{2}{c}{\text{DV's}} & \text{TV's} & & \text{Errors} \\
\text{DV's} & Y_1 & Y_2 & X_1 \quad X_2 \quad X_3 & & \in \\
Y_1 & 0 & 0 & \gamma_{11}X_1 + \gamma_{12}X_2 + \gamma_{13}X_3 & + & \in_1 \\
Y_2 & \beta_{21}\gamma_1 & 0 & \gamma_{21}X_1 + \gamma_{22}X_2 + \gamma_{23}X_3 & + & \in_2
\end{array}
$$

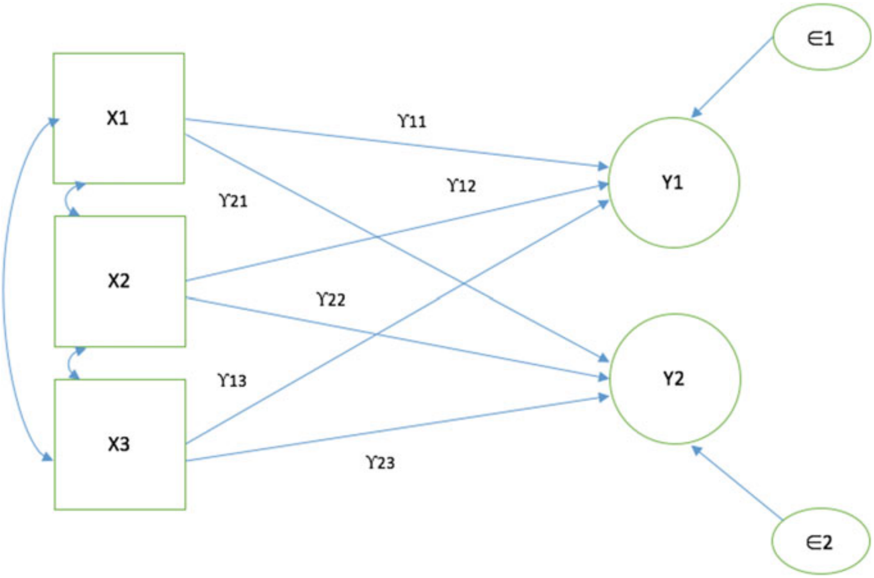


Fig. 2.1 General path model

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + \begin{bmatrix} \in_1 \\ \in_2 \end{bmatrix}$$

$$\gamma = \beta\gamma + \Gamma X + \in \tag{2.1}$$

$$\begin{aligned} (l-\beta)_\gamma &= \Gamma X + \in \\ (I-\beta)^{-1}(l-\beta)\gamma &= (1-\beta)^{-1}\Gamma X + \in \\ \gamma &= (I-\beta)^{-1}\Gamma X + \in \end{aligned} \tag{2.2}$$

2.5 Confirmatory Factor Analysis (CFA)

The measurement of latent variables is known as confirmatory factor analysis. It extracts the latent construct from other variables and shares the most variance with related variables. It evaluates latent variables on the basis of causal or correlated variation of the dataset and decreases the dimensions of data, can standardize the scale for several indicators and justify for the correlations present in dataset. Hence, it is necessary to understand the reason for the use of latent variable before it is incorporated into the model. There are two factor analysis techniques, one is CFA and EFA (exploratory factor analysis). EFA is used when we need to investigate

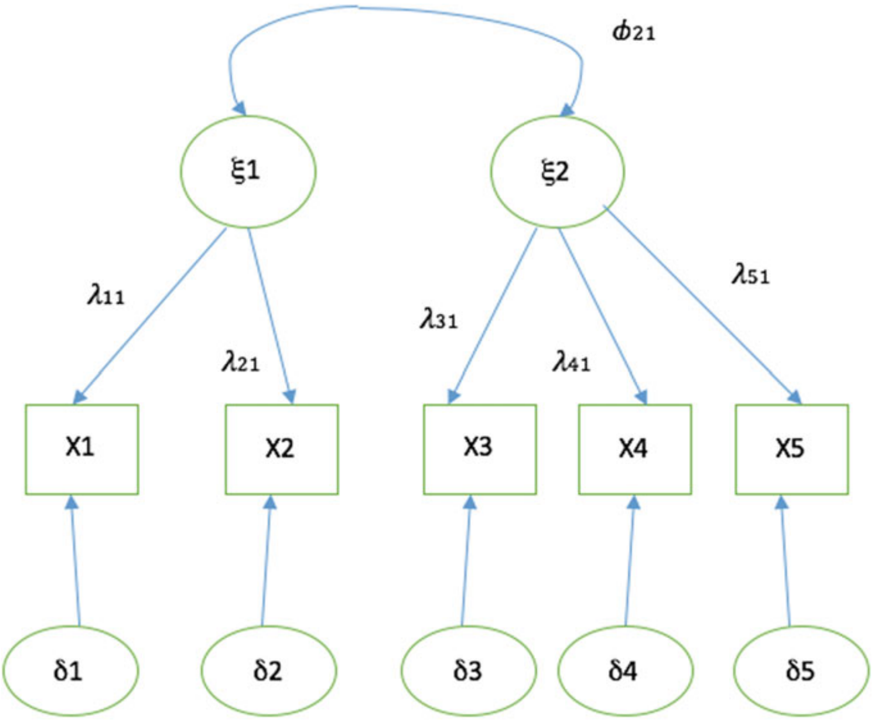


Fig. 2.2 General CFA model

the underlying latent variables, whereas CFA is used when the indicators are well specified according to either related theories or prior knowledge (Fig. 2.2).

X_1, X_2, X_3, X_4, X_5 are the measured variable.

In the measurement model, common notations used are as follows

$$\begin{aligned} \xi &= \text{Latent Factors} \\ X &= \text{Measured Variables\%} \\ \lambda_x &= \text{Factor Loadings} \\ \delta_x &= \text{Errors} \\ \Phi &= \text{Correlation between the constructs} \end{aligned}$$

Equations (2.3)–(2.7) represent the directional influences of the exogenous LVs (ξ) on their indicators (x). Equation (2.8) links the observed (manifest) variables to unobserved (latent) variables through a factor analytic model and constitutes the "measurement" portion of the model. Equation (2.9) is final measurement equation.

$$X_1 = \lambda_{11}\xi_1 + \delta_1 \tag{2.3}$$

$$X_2 = \lambda_{11}\xi_1 + \delta_2 \tag{2.4}$$

$$X_3 = \lambda_{11}\xi_2 + \delta_3 \tag{2.5}$$

$$X_4 = \lambda_{42}\xi_2 + \delta_4 \tag{2.6}$$

$$X_5 = \lambda_{52}\xi_2 + \delta_5 \tag{2.7}$$

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ 0 & \lambda_{32} \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{bmatrix} \tag{2.8}$$

Measurement Equation

$$X = \Lambda\xi + \delta \tag{2.9}$$

2.6 Mathematical Specification of Structural Equation Modelling

A structural equation model can be defined as a hypothesis of a specific pattern of relations among a set of measured variables (MVs) and latent variables (LVs). The three equations presented below are fundamental to SEM. Equation (2.10) represents the directional influences of the exogenous LVs (ξ) on their indicators (x). Equation (2.11) represents the directional influences of the endogenous LVs (η) on their indicators (y). Equations (2.10) and (2.11) link the observed (manifest) variables to unobserved (latent) variables through a factor analytic model and constitute the "measurement" portion of the model. Equation (2.12) represents the endogenous LVs (η) as linear functions of other exogenous LVs (ξ) and endogenous LVs plus residual terms (ζ). Thus, Eq. (2.12) specifies relationships between LVs through a structural equation model and constitutes the "structural" portion of the model (Fig. 2.3).

Here,

$\eta = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}$ = Endogenous Variable (Dependent Variable)

$\zeta = \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$ = Error Term Related to η

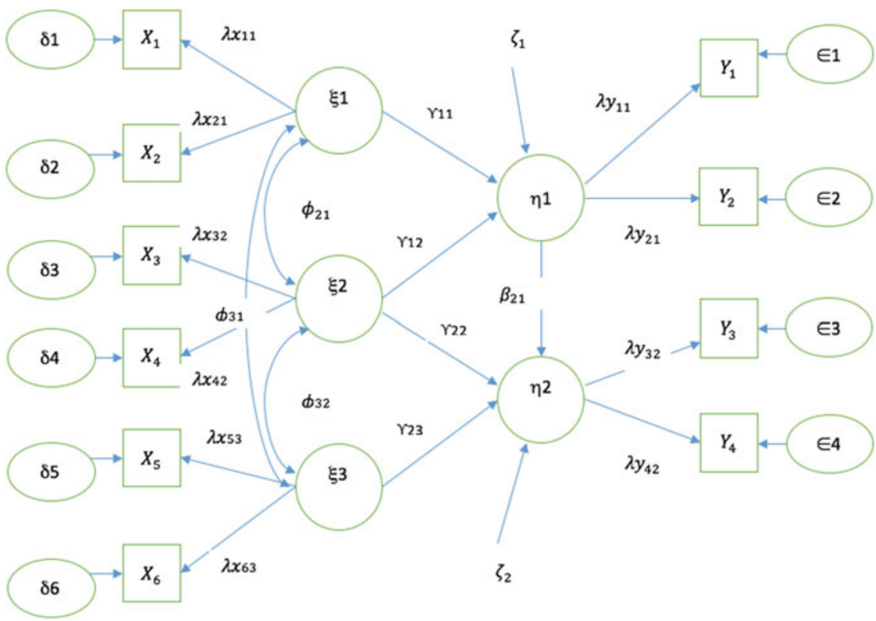


Fig. 2.3 General SEM model

$$\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} = \text{Exogenous Variable (Independent Variable)}$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \text{Dependent Manifest Variable}$$

$$\in = \begin{bmatrix} \in_1 \\ \in_2 \\ \in_3 \\ \in_4 \end{bmatrix} = \text{Error Term Related to } Y$$

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \text{Independent Manifest Variable}$$

$$\delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix} = \text{Error Term Related to } X$$

$$\varphi = \mathrm{Cov}(\xi)$$

$$\psi = \mathrm{Cov}(\zeta)$$

It can also be written in matrix format as follows

$$
\begin{array}{cccc}
 & \eta & \xi & \text{Errors} \\
\eta & \eta_1 \quad \eta_2 & \xi_1 \quad \xi_2 \quad \xi_3 & \zeta \\
\eta_1 & 0 \quad 0 & \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + 0 \quad + & \zeta_1 \\
\eta_2 & \beta_{21} \quad 0 & 0 + \gamma_{22}\xi_2 + \gamma_{23}\xi_3 \quad + & \zeta_2
\end{array}
$$

$$
\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & 0 \\ 0 & \gamma_{22} & \gamma_{23} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}
$$

Structural Equations

$$
\begin{aligned}
\eta_1 &= \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \zeta_1 \\
\eta_2 &= \gamma_{22}\xi_2 + \gamma_{23}\xi_3 + \zeta_2 \\
\eta &= \Gamma\xi + \beta\eta + \zeta
\end{aligned}
$$

Measurement Equations

$$
\begin{aligned}
X &= \Lambda_x \xi + \delta \\
y &= \Lambda_y \eta + \varepsilon
\end{aligned}
$$

Hence, to conclude

$$x = \Lambda_x \xi + \delta \tag{2.10}$$

$$y = \Lambda_y \eta + \varepsilon \tag{2.11}$$

$$\eta = B\eta + \Gamma\xi + \zeta \tag{2.12}$$

where δ is the measures of exogenous manifest variables, Λ_x the effect of exogenous LVs on their MVs (matrix), d the error of measurement in exogenous manifest variables, y the measures of endogenous manifest variables, Λ_y the effect of endogenous LVs on their MVs (matrix), ε the error of measurement in endogenous manifest variables, ξ the latent exogenous constructs, η the latent endogenous constructs, Γ the effect of exogenous constructs on endogenous constructs (matrix), B the effect of endogenous constructs on each of the other endogenous constructs (matrix) and ζ is the errors in equations or residuals. It indicates the difference between the actual and estimated value of any relationship in the model. It is basically the difference between the observed and estimated fitted covariance matrices in SEM analysis.

It is also necessary to define the following covariance matrices:

(a) $\Phi = E(\xi\xi')$ is a covariance matrix for the exogenous LVs.
(b) $\theta_\delta = E(\delta\delta')$ is a covariance matrix for the measurement errors in the exogenous MVs.
(c) $\theta_\varepsilon = E(\varepsilon\varepsilon')$ is a covariance matrix for the measurement errors in the endogenous MVs.
(d) $\psi = E(\zeta\zeta')$ is a covariance matrix for the errors in equation for the endogenous LVs.

From this mathematical representation, it can be deduced that the population covariance matrix for the MVs is a function of eight parameter matrices: Λ_x, Λ_y, Γ, B, Φ, θ_δ, θ_ε and ψ. Once we have a hypothesized model including fixed and free parameters of the eight parameter matrices, and a sample covariance matrix for the MVs, it is possible to solve for estimates of the free parameters of the model. The most common approach for fitting the model to data is to obtain maximum likelihood estimates of parameters and an accompanying likelihood ratio χ^2 test of the null hypothesis that the model holds in the population.

2.7 Summarization of Key Concepts

The first task of the researcher is to identify and define the series of relationships to identify a suitable model for analysis. The constructs are defined either as exogenous or endogenous. This is followed by the development of path diagram which provides a graphical representation of the relationships. It shows impact of one construct on another using straight line arrow. If research does not anticipate any relationship between two constructs, then there will be no path (arrow) between these two constructs. If a researcher has hypothesized a causal relationship in the model, then such dependence relationships are represented with arrows pointing from the cause to the consequent effect. Once the researcher has specified relationships and path diagram, he can focus on data collection in the suitable format. An analysis of this data in SEM helps to estimate the strength of the relationships and examine how well the data actually fits the model.

There is a fundamental difference in the measurement model and structural model. The model which relates measured variables to latent variables is called measurement model. The structural model relates latent variables to one another. The purpose of measurement model is to examine the relationship between the latent variables and their measures. The structural model evaluates the relationships between latent variables. To test the measurement model, we typically saturate the structural model, by allowing all the latent variables to correlate.

A widely used application of SEM is a confirmatory modelling strategy. Here, the objective of the researcher is to investigate how well the model fits the data. In this case, if a researcher gets an acceptable fit of the proposed models, then it is called the confirmation of the proposed model, and one can say that the proposed model is one of the several possible acceptable models which provide a good fit to the data. This process does not prove the proposed model, but the purpose here is just to confirm it (Fig. 2.4).

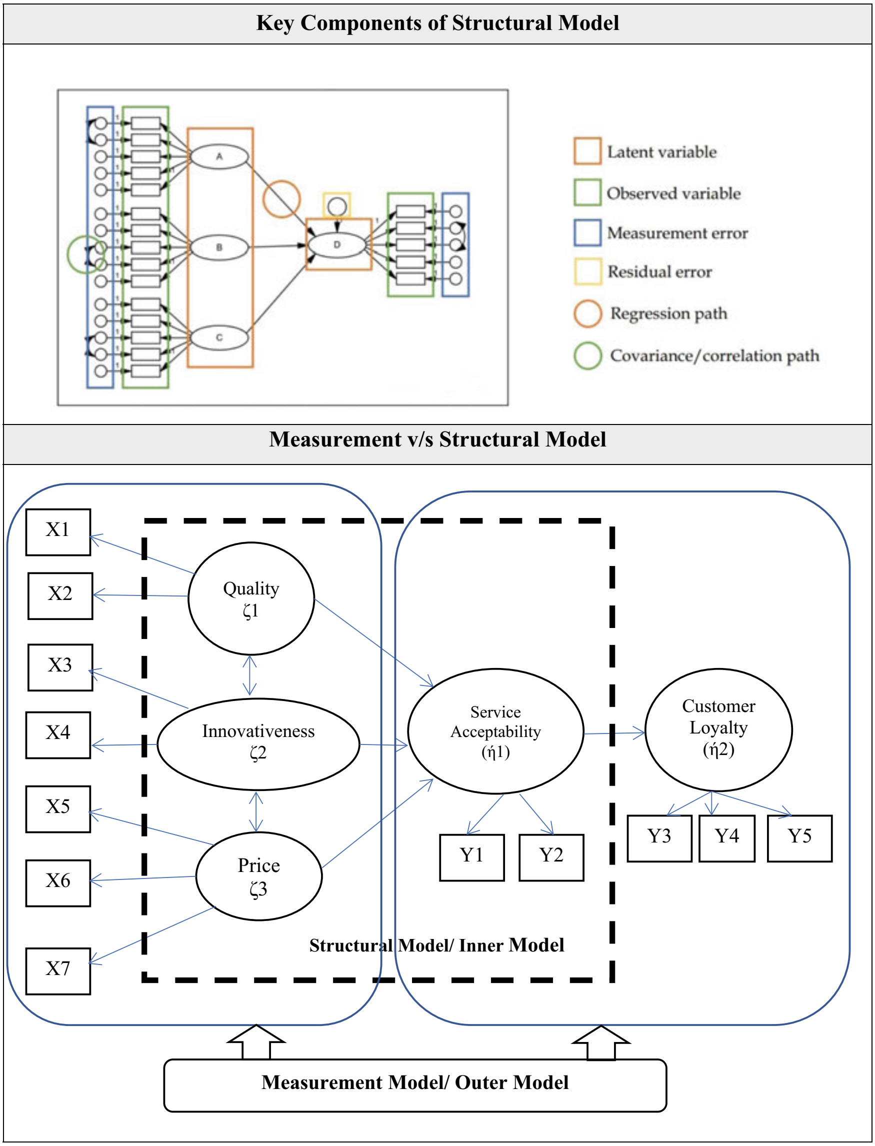


Fig. 2.4 Measurement versus structural model

2.8 Importance of SEM in Research

Any research question including direct and indirect relationship between dependent or independent variables can be solved with the help of SEM. The first and foremost goal of SEM is to determine and validate the anticipated causal model or process. Hence, SEM is a confirmatory model or technique. It is similar to any other test, as we have a sample and then there is an assumption about the population that includes the sample. On the basis of sample, we have the covariance matrix which serves as the dataset. The proposed model should produce a population covariance matrix which should be consistent with the sample covariance matrix, and hence, one must specify a priori model that will undergo a validation testing.

Structural equation model can let you know that if the model is adequate or not. In SEM, parameters are evaluated and equated with the sample covariance matrix. For adequacy, goodness of fit is calculated which will let you know the model is appropriate or not. It also lets you know about the amount of variance in both manifest and latent variables (dependent variables) is because of independent variables. SEM also tells the reliability of each measured variable. SEM also examines mediation and moderation. SEM can also reveal about group differences by fitting separate SEM into different groups and hence compare results. Also, both random and fixed effects can be included in the model and hence enable us to use hierarchical modelling techniques in analyses.

2.9 Assumptions in SEM

Structural equation modelling is also called causal modelling because it tests the proposed causal relationships. The following assumptions are assumed:

1. **Multivariate normal distribution**: The maximum likelihood technique is used and assumed for multivariate normal distribution. Small changes in multivariate normality can lead to a large difference in the chi-square test.
2. **Linearity**: Endogenous variables and exogenous variables have linear relationship.
3. **Outlier**: It affects the model significance and hence it is necessary to ensure that the data is free of outliers.
4. **Sequence**: The cause should occur before the event. Endogenous and exogenous variables should have an effect and cause relationship.
5. **Non-spurious relationship**: Covariance observed should be correct.
6. **Model identification**: In SEM, underidentified models are not being considered, and equations should be greater than the estimated models should be exactly identified.
7. **Sample size**: As the rule of thumb, sample size is generally 20 times more than the indicator. For instance, mostly researcher prefers 150–300 sample size with 10–15 indicators.

8. **Uncorrelated error terms**: All the error terms have the assumption of being uncorrelated with other variable errors.
9. **Data**: SEM uses interval data set.

2.10 Sample Size Considerations in SEM

In statistical methods, sample size provides a basis for estimating the sampling error. As it is difficult and expensive to deal with larger samples, the critical question in SEM analysis like other statistical method is to address a question—"what should be a reasonable sample size to produce acceptable and trustworthy results". Typically, sample size in SEM is governed by five issues: (i) multivariate distribution of the data; (ii) estimation technique used; (iii) underlying complexity of the model; (iv) amount of missing data; (v) amount of average error variance among the reflective indicators.

Sample size considered in SEM is huge. SEM is capable of handling complex relationship among multivariate data, and thus, sample size is important. The two popular assumptions made in case of sample size are that you should have at least 50 or more than 200 observations more than eight times the number of variable used model.

The sixth edition of Hair et al. (2012) provides a systematic and easy to implement guidelines on sample size selection. The recommendations include

	Conditions/considerations in SEM model	Recommended sample size
Recommendation 1	Five or fewer constructs, each with more than three items (observed variables) and with high item communalities (0.6 or higher)	100–150
Recommendation 2	Modest communalities (0.45–0.55) or the model contains constructs with fewer than three items	200
Recommendation 3	Lower communalities with multiple underidentified (fewer than three items) constructs	More than 300
Recommendation 4	More than six factors with select factors using less than three measured items as indicators Presence of multiple low communalities	More than 500

2.11 Key Issues in SEM

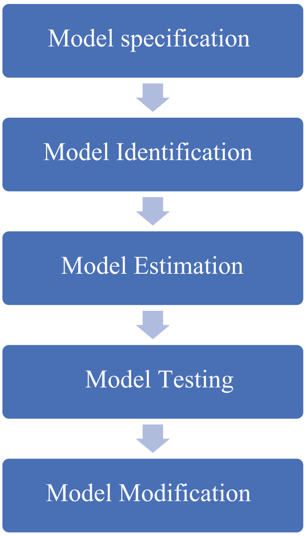
- SEM is a confirmatory technique; hence, full model should be specified and tested depending on samples and variables used in calculations. Number of parameters should be known beforehand such as variance, covariance and path coefficients. Also, all the relationships used in the model should be known, and then, only one could begin the analysis.
- Multivariate normality is required in SEM analysis. Univariate and multivariate outliers should be examined. Like other multivariate statistical methodologies, most of the estimation techniques used in SEM require multivariate normality.
- First-order (linear) relationships among variables are only considered in SEM technique. Bivariate scatter plots are created to explore linear relationships for the variables. If the relationship among two variables is quadratic, the power transformations can be done.
- Identifying multicollinearity between the independent variables is a complex issue. Usually, most of the procedures inspect the determinant of a section of your covariance matrix or the whole covariance matrix. A very small determinant may be indicative of extreme multicollinearity.

Chapter 3
Procedural Steps in Structural Equation Modelling

Structural equation modelling includes six key steps. In addition to data collection, the steps are model specification, identification, estimation, evaluation and modification (Fig. 3.1).

1. **Model Specification**: The first step in SEM analysis is the model specification. It is performed prior to data collection and data modelling. This involves the development of a theoretical model which defines the variables and their relationships based on the existing literature and theory. This process is difficult, and hence, it is advised that the model must be grounded and derived from the existing literature. The model should be well defined, and researcher should be able to explain the relationship in the model and the rationale of the overall model. The first step includes that the measurement model is specified with all the latent constructs.

Fig. 3.1 Steps in SEM analysis



J. J. Thakkar, *Structural Equation Modelling*, Studies in Systems, Decision and Control 285, https://doi.org/10.1007/978-981-15-3793-6_3

The structural model is specified when the latent construct in the measurement model is aptly measured by the observed variable as measurement model does not specify the directional relationship between the latent variables. The structural model specifies the relationship between latent variables in theoretical form. It is obvious that such relationship should be specified before estimation and testing of the model. The structural equation obtained evaluates the specific structure coefficient. Each equation has a prediction error which specifies the degree of variance in the latent endogenous variables. The equation also specifies the predicted relationships. These relationships between latent and observed variables are also shown by the path diagram.

2. **Model Identification**: This is the second step in SEM analysis and happens prior to the estimation of model parameters which is the relationship between the variables in the model. Model identification is concerned with the task of whether the unique solution can be formed for the model or not. It must theoretically establish unique estimate for each variable for the model to be identified. Model identification is dependent on the parameters as free, fixed and constrained. Free parameters are those parameters which are unknown and which need to be estimated. Fixed parameters are those which are fixed at a specific value such as 0 or 1. Constrained parameters are those which are unknown but constrained to one or more other parameters. To identify the structural equation model, measurement model must be identified. The measurement model is identified in these two conditions. Firstly, there are two or more latent variables, each has at least three indicators loaded on it, and the errors of these variables are not correlated. It is also necessary that each indicator loads only one factor. Secondly, there are two or more latent variable, each has only two indicators loaded on it, and the error of these variables is not correlated. It is also necessary that each indicator loads only one factor, and variance or covariance among these factors is zero. A casual path from each latent variable to the observed variable should be zero for the likelihood of identification. Hence, reference variable is the variable which is fixed and non-zero loaded and has the most reliable scores.

A structural model to be identified could be extremely cumbersome and involves a highly complex calculation, and hence, structural model has to be outlined with a set of rules such as the recursive rule and *t*-rule. A structural model is said to be recursive when the model is unidirectional, i.e. when the two variables are related in only one directional. According to *t*-rule, the equation should have known pieces of information more than unknown pieces of information to determine a unique solution.

3. **Model Estimation**: The third step of the analysis is called model estimation. It estimates the theoretical model parameters in a way that the theoretical parameter values give a covariance matrix close to observed covariance matrix S. Structural modelling equation uses a iterative feature also known as fitting function. Fitting function is used to minimize the difference between the observed covariance matrix S and the estimated theoretical covariance matrix P, and hence, it improves

the primary estimates of parameter with iterative calculation cycle. The final estimates give the best fit parameter to the observed covariance matrix S. Several estimation procedures are available such as least squares, maximum likelihood, asymptotic distribution free (ADF), unweighted least squares and generalized least squares. Maximum likelihood (ML) is most commonly used estimation technique which is followed by generalized least squares (GLS). Although ML and GLS are comparable to ordinary least squares (OLS) estimation used in multiple regression, they possess select key advantages over OLS estimation. In particular, ML and GLS are (a) not scale-dependent, (b) allow dichotomous exogenous variables and (c) offer consistent and asymptotically efficient results in large samples. ML and GLS assume multivariate normality of dependent variables and are called full information techniques as they estimate all model parameters simultaneously to produce a full estimation model. This is a key limitation with OLS. A use of asymptotically distribution-free (ADF) estimator is recommended when the assumption of multivariate normality is violated. ADF does not depend on the underlying distribution of the data, but it requires a large sample size as the estimator yields inaccurate chi-square (χ^2) statistics for smaller sample sizes.

4. **Model Testing**: Structural equation modelling helps in concurrent analysis of both indirect and direct relationship among manifest and latent variables. The model testing includes the analysis of two conceptually distinct models such as structural and the measurement model. It is necessary for a researcher to ensure that the observed variable chosen for the latent variable is actual measure of construct. In absence of such verification, the structural model becomes meaningless. Model fitting has a problem that power varies with the sample size. For example, if we have a very huge sample size, then the sample test will always be significant, and hence, we reject the model, and on the other side, if we consider a very small sample, then model will always be accepted though it fits badly.

Model evaluation indices

Evaluation of SEM is generally based on fit indices to test the single path coefficient such as p value and standard error and for the overall model such as χ^2, RMSEA. There are three types of model-of-fit indices:

- Absolute fit indices
- Incremental fit indices
- Parsimony fit indices (Fig. 3.2).

Absolute fit indices

- It measures the overall goodness of fit for both the structural and measurement models collectively.
- Absolute fit indices are the indices which show that how well a priori model will be fitting in the sample data. It also shows that which model has the best superiority among the proposed models.

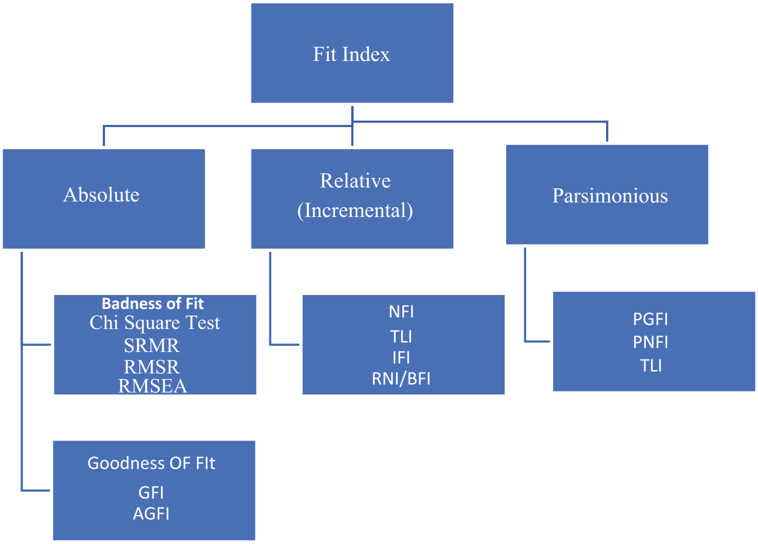


Fig. 3.2 Types of fit index

- The indices include the following: Models chi-square (χ^2), relative/normed chi-square (χ^2/df), RMSEA, GFI, AGFI, RMR and SRMR.

Incremental fit indices

- It is also known as comparative indices or relative fit indices.
- The indices include the following: NFI, NNFI and CFI.

Parsimony fit indices

- Estimation process is dependent on data when we have nearly saturated or complex model. Hence, we need less rigorous theoretical model that paradoxically gives better fit indices.
- The indices include the following: PGFI and PNFI (Table 3.1).

It is difficult to specify a generic guideline for fitness indices which can help the researcher to distinguish a good model from a poor model. However, the selected recommendations are outlined by Hair et al. (2012) include

- Goodness of the model should be verified using three to four indices of different types.
- Index cutoff values should be adjusted based on model characteristics.
- Employ the use of multiple indices to examine the goodness of the model. This helps the researcher to determine which model is better when the set of acceptable models are compared with the help of multiple indices.

Table 3.1 Classification of fit indices and their cutoff value

Model-of-fit indices	Full name/key concerns	Cutoff value
Absolute fit indices		
Model chi-square $\left(\chi^2\right)$	Chi-square (use only for sample $n < 200$ or $p > 0.05$)	Insignificant result ($p > 0.05$)
χ^2/df	Relative/normed chi-square (use only for sample $n > 200$ or if $p < 0.05$)	Value of <2.0
RMSEA	Root mean square error of approximation	Value between 0.08 and 0.10 (mediocre fit), <0.08 (goof fit)
GFI	Goodness of fit statistics Exhibits bias towards samples	Value >0.90 or >0.95 (use 0.95 if factor loading and number of sample are low)
AGFI	Adjusted goodness of fit statistics Needs to be accompanied by other indices	Value of >0.80
RMR	Root mean square residual	N/A
SRMR	Standardized root mean square residual	Value of <0.05
Incremental fit indices		
NFI	Normed fit index Sensitive to sample size <200 Must be accompanied by other indices	Value of >0.90
NNFI	Non-normed fit index	Value of >0.80
CFI	Comparative fit index Revised version of NFI Less affected by sample size	Value of $\geq$0.90
Parsimony fit indices		
PGFI	Parsimony goodness of fit index	Value of >0.90
PNFI	Parsimonious normed fit index	Value of >0.90

5. **Model Modification**: This is the final step of structural equation modelling. A researcher intends to modify the model so that they explore the best-fitted model which fits the data perfectly. Firstly, researcher has to accomplish a model specification search which eliminates the non-significant parameters from the theoretical model which is also known as theory trimming, and then, they need to examine the model's standardized residual matrix which is called as fitted residuals. To eliminate parameters, one common method is to compare the *t*-static for individual parameter to the tabled *t*-value to find its significance in the sample. While examining the standardized residual matrix, one should attempt to find all the values which are small in magnitude as large values in the matrix imply a misspecification of the general model, whereas large values across an individual variable imply to the misspecification in that individual variable only.

The above-said procedures can improve the fit of the model, but it is highly contradicted method as specification searches are exploratory in nature and thus are based on the sample data instead of previous theory and research, and as a result, parameter eliminated from the theoretical model may reflect sample characteristics that do not generalize to the broader population. Also, model modification may progress to an inflation of Type I error, and hence, it can be misleading. Therefore, it is recommended that a researcher should keep a balance while eliminating the parameters in the model to improve the fit of the model.

In summary, there are six steps in conducting structural equation modelling.

Step 1: Define the individual constructs. Typically, this should answer "what items to be used as measured varaibles?"
Step 2: Develop and specify the measurement model. This includes two things: (i) associate measured variables with constructs and (ii) develop a path diagram for the measurement model.
Step 3: Design a study to produce empirical results. Here, a researcher must examine the adequacy of the sample size. In addition, he should select an appropriate estimation method and missing data approach.
Step 4: Assess the validty of measurement model. This should be done by examining goodness of fit (GOF) indices and construct validity of measurement model. If the mesurement model is valid as per the prescribed ranges of GOF and construct validity then a researcher can move to step 5. If this is not satisfied then a researcher must refine the measures and design a new study.
Step 5: Specify structural model. This requires measurement model to be converted into structural model by assigning relationships from one constructs to another based on the proposed theoretical model.
Step 6: Finally, a researcher should assess the validity of structural model by checking goodness of fit indices (GOF) and signficance, direction and size of structural parameters estimates. If the structural model is valid then researcher can draw susbstantive conclusions and extend necessary recommendations. If structural model fails the test of validity then a researcher should refine the model and test with new data set.

Chapter 4
Applications of Structural Equation Modelling with AMOS 21, IBM SPSS

4.1 History of AMOS, IBM SPSS Software

AMOS is a highly popular software with a unique graphical user interface (GUI) for solving structural equation modelling problems. The software is developed by IBM and SPSS Inc. Before 2003, AMOS software was part of SmallWaters Corp. AMOS is extensively used by the researchers for multivariate analysis by integrating the use of various multivariate analysis methods such as regression, factor analysis, correlation, and analysis of variance. AMOS provides an intuitive graphical or programmatic user interface for evaluating the complex relationships among the constructs. SPSS AMOS is available for the Windows operating system. The company SPSS Inc. was started in 1968 with a headquarter in Chicago for the development of proprietary software of the name SPSS. In 2009, the company was acquired by IBM and it has become IBM SPSS which is now fully integrated into the IBM Corporation and is recognized under IBM Software Group's Business Analytics Portfolio. The software is widely used by researchers in the domain of social sciences and management for the following reasons:

- It has an intuitive drop-down menu which is easy to learn for beginners.
- It offers much flexibility and customization in writing syntax.
- It is supported on Windows operating system.

SPSS AMOS is licenced software, and an authentic version needs to be purchased from IBM SPSS. Company offers a 30-days free trial version to explore the capabilities of this software. It is necessary to renew and upgrade SPSS AMOS licence yearly.

J. J. Thakkar, *Structural Equation Modelling*, Studies in Systems, Decision and Control 285, https://doi.org/10.1007/978-981-15-3793-6_4

4.2 A Step-by-Step Procedure to Solve SEM Using AMOS, IBM SPSS

Structural equation modelling (SEM) is a technique for observing interdependencies among the various variables. It is a confirmatory method to check for the fitting of data and conceptual models. It is mainly used in the areas of drawing conclusions wherever non-quantifiable variables are present. For example, measuring satisfaction or happiness. These variables cannot be measured directly and are measured using a survey which includes various questions which act as identifiers or indicators for the variable—happiness. For example, satisfaction and peace of mind are some of the questionnaire questions to check the happiness of a person. SEM is also called as the combination of factor analysis and multiple regression. This can be solved using IBM SPSS AMOS, LISREL, R, etc., if solving on a software platform.

The basic statistics of the SEM are the covariance, variance, correlations and regression coefficients. Before starting the structural equation modelling and forming the conceptual model, the following terms have to be understood.

1. Latent variables: These are intangible which are indirectly measured using identifiers or measured variables. It is denoted by an ellipse.
2. Structural model: This is a diagram which indicates the overall concept of the model using circles, rectangles, ellipses and arrows.
3. Measured variables: These are the variables which can be measured and act as identifiers for the latent variables. These are denoted by rectangles.

The procedure for conducting a structural equation modelling analysis using a software tool is explained using an example model of job satisfaction of the employees and its relationship between the human resource policies in a company.

The following steps have to be followed in order to conduct SEM analysis in AMOS software.

Step 1: Develop a conceptual model and hypothesis for the problem statement
Figure 4.1 shows the intention of the problem and its causes as well. In this example, the job satisfaction is dependent on flexibility in the work and the talent development procedure by the HR team. The hypothesis formulated by the modeller is as follows:

H1: High flexibility has high job satisfaction.
H2: High talent development process has high job satisfaction.

Step 2: Develop suitable identifiers and collect data regarding the same
It can be noticed that these three variables shown in Fig. 4.1 cannot be quantified. Hence, there has to be many other indirect measurable variables which can denote the satisfaction level, flexibility level and the development of the employee. It is left to the modeller to decide the measurable variables and can be found by doing a literature review related to the problems as per the conceptual model. Here, the measurable variables for the job satisfaction can be identified by the variables—extent of productivity (S1), the commitment of employee (S2) and the punctuality of the

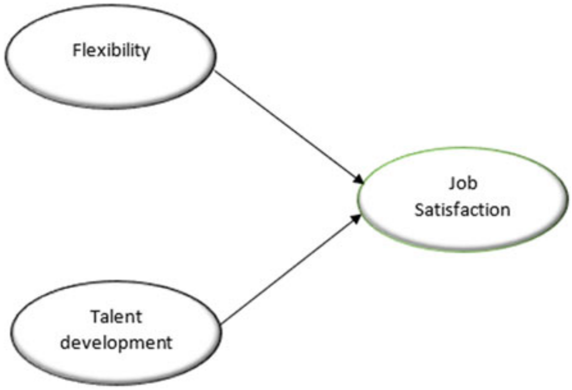


Fig. 4.1 Conceptual model of the problem statement

employee (S3). These variables indirectly denote the satisfaction of the employee and are called identifiers. The identifiers are quantified using a questionnaire with a rating scale from 1 to 5 (any scale having a minimum and maximum value) indicating a negative response for 1 and a positive response for 5. This can be collected and entered in a spreadsheet and can be saved in few compatible formats which are acceptable by the software. Few of the acceptable formats by AMOS are ***.csv, .txt, .sav (IBM SPSS Statistics is required for this format), .xls (older than MS Excel 8), .AMOSRecode,*** etc. Similarly, identify the identifiers for the other two latent variables—flexibility and talent development. The identifiers for the talent development are—training offered (T1), promotions in the post (T2) and salary (T3). Similarly, for flexibility, the location of the company (F1) and work hours (F2) are the identifiers. The questionnaire consisting of 100 responses recorded in Excel is shown in Fig. 4.2. The rating is sorted from highest score to lowest score, and this sorting does not matter.

Step 3: Developing a model using proper notations

In a structural equation model, a single-headed arrow is used to connect the effector to the effected variables. When there are only single-headed arrows originating from a variable, the variable is called exogenous and if any single-headed arrow is terminating or ending at a variable, then it is called endogenous. There exists always a covariance between two exogenous variables, and it has to be connected by double-headed arrows. The endogenous variables and the measurable variables always have an error in the estimation and have to be indicated by a circle wherein a single-headed arrow leads from the error to the variable. This is explained in the further steps as and when the software applicability is introduced.

Step 4: Drawing model in IBM SPSS AMOS

Open the ***AMOS Graphics*** application in the computer, and the screen looks as shown in Fig. 4.3. The paper size on the screen can be changed by going to ***interface properties*** available in the ***view*** tab (marked in red). The left-hand side of the screen provides all the tools to create a model and analyse it. The ***view*** option can also be used to select the required output from the analysis.

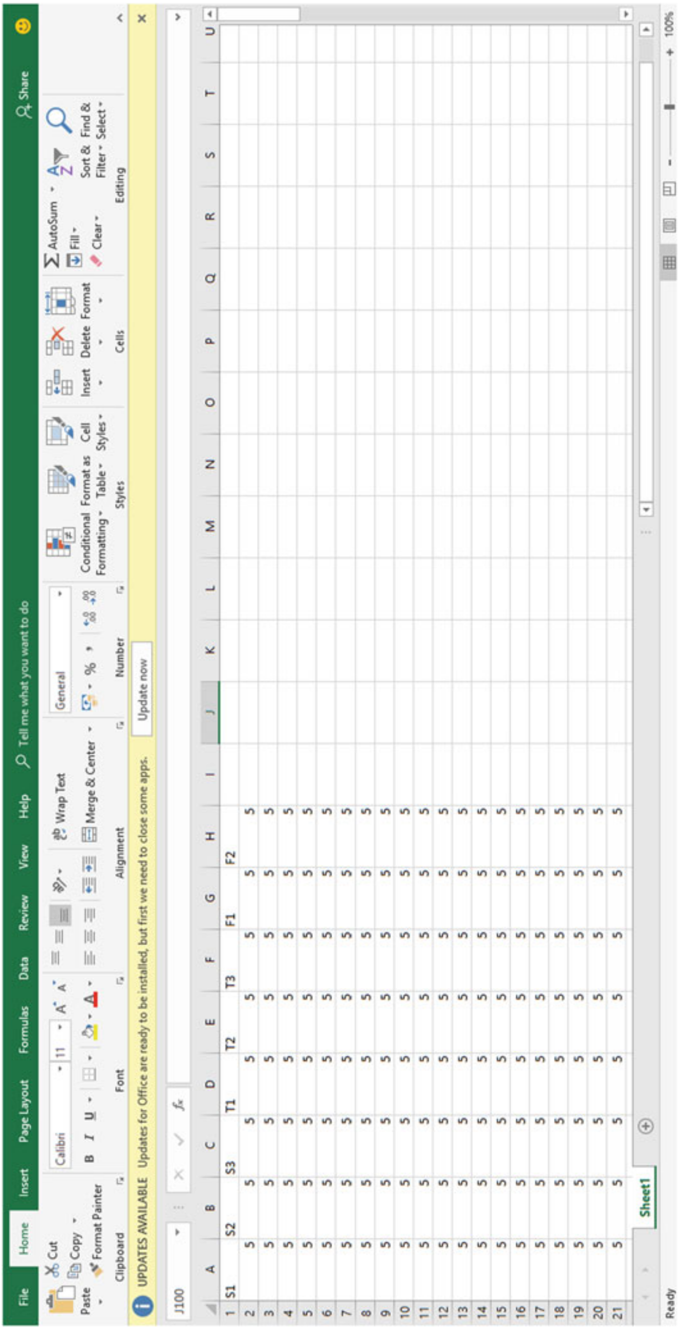


Fig. 4.2 Data set from the questionnaire in MS Excel

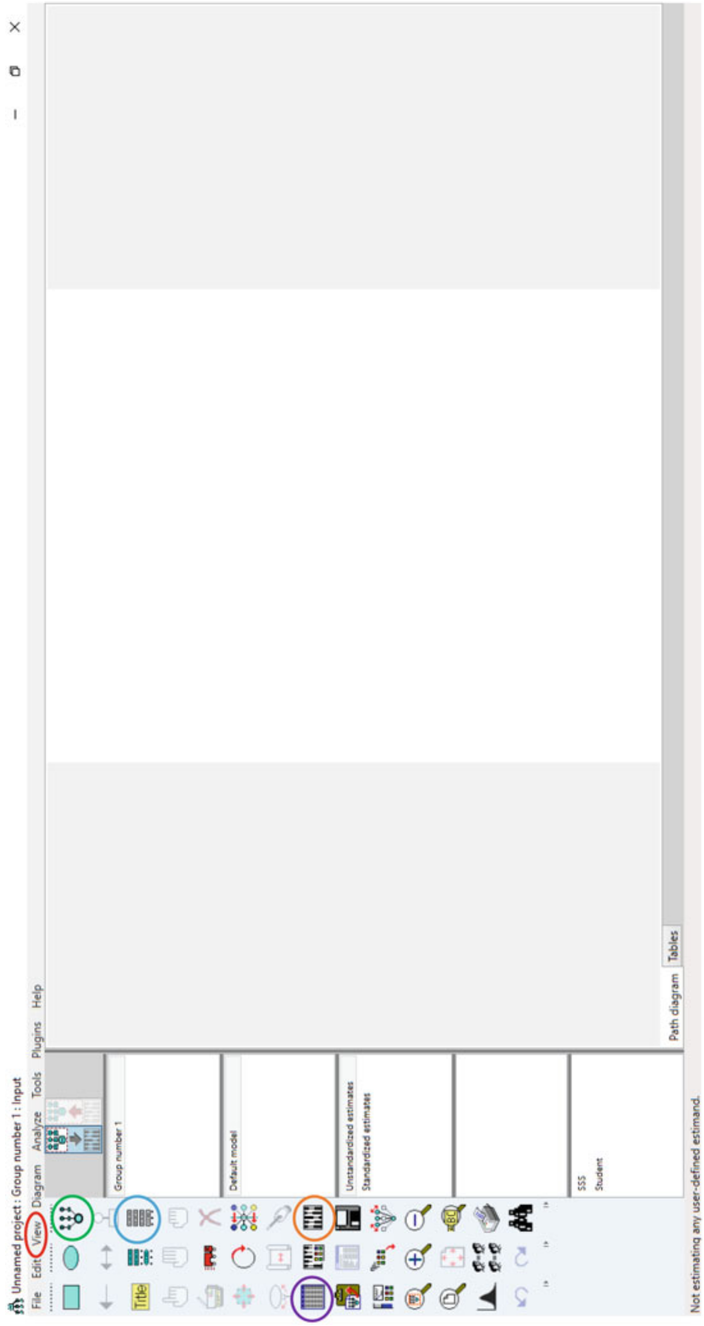


Fig. 4.3 Home screen of IBM SPSS AMOS 21

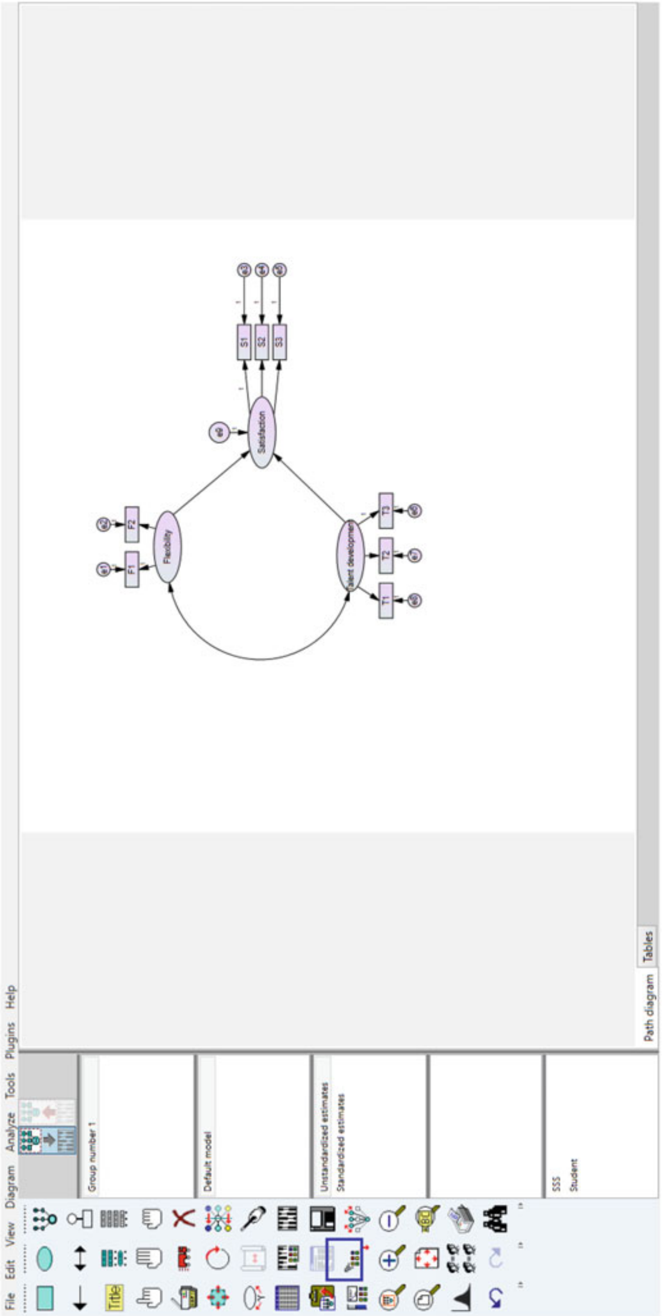


Fig. 4.4 AMOS screen after drawing the model

Referring to Fig. 4.3, select the icon (marked in green) to draw a latent variable. To draw the identifiers and its errors, click on the latent variable figure (ellipse). Rename any variable by double-clicking on the figure. Later, use the arrow marks which is in the second row of the tools. In the current problem, use the tool marked in green and click on the sheet three times at different places to get the latent variables. Double-click each variable and rename them as job satisfaction, flexibility and talent development. To add the identifiers and errors, use the same icon (green) and click on the latent variables as many times as the number of identifiers. Use single-headed arrows and connect between talent development and job satisfaction, also connect the flexibility with job satisfaction. Using double-headed arrows, connect the exogenous variables—talent development and flexibility. The variables' shapes can be rotated by the symbol under the ***red truck***. The ***red truck*** can be used to move the variables anywhere on the screen.

Step 5: Adding the data to the identifiers

Before adding the data, the file containing the data has to be selected. Use the icon (marked in purple) to select the data file. After clicking the ***select data*** icon, a pop-up window is displayed as shown in Fig. 4.5. Select a data file by clicking on ***file name***. In order to allocate the data set, click on the ***list variables in the data set*** icon (marked in blue) and a pop-up will be displayed as shown in Fig. 4.6. Drag the variables to the required identifier figures in the model (rectangles). The rectangles will be automatically renamed as per the variable. In order to name the unnamed variables, click on ***name unobserved variables*** from the ***plug-ins*** tab available at the top of the screen. Also, add an error variable for the job satisfaction using the icon below the icon marked in green. After all the things are completed, the model looks as shown in Fig. 4.4.

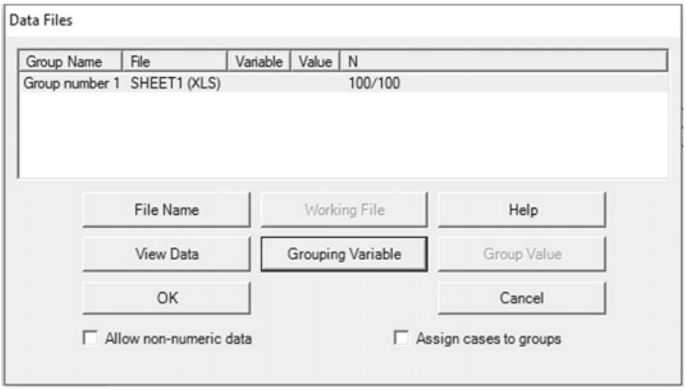


Fig. 4.5 Pop-up window for selecting a data file

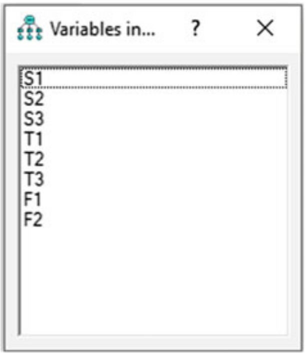


Fig. 4.6 Pop-up window to add the data set to the variables

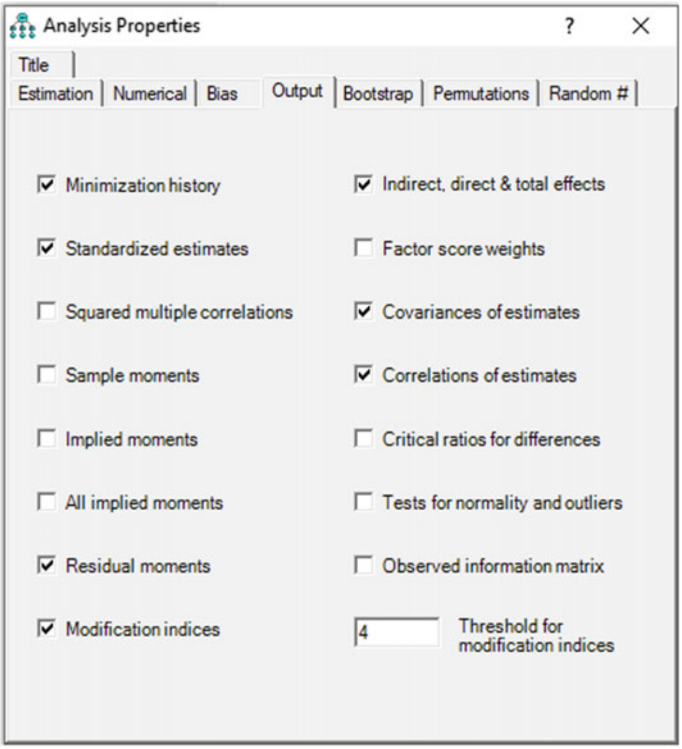


Fig. 4.7 Analysis properties pop-up window from the view tab

Step 6: Solving and fitting the model

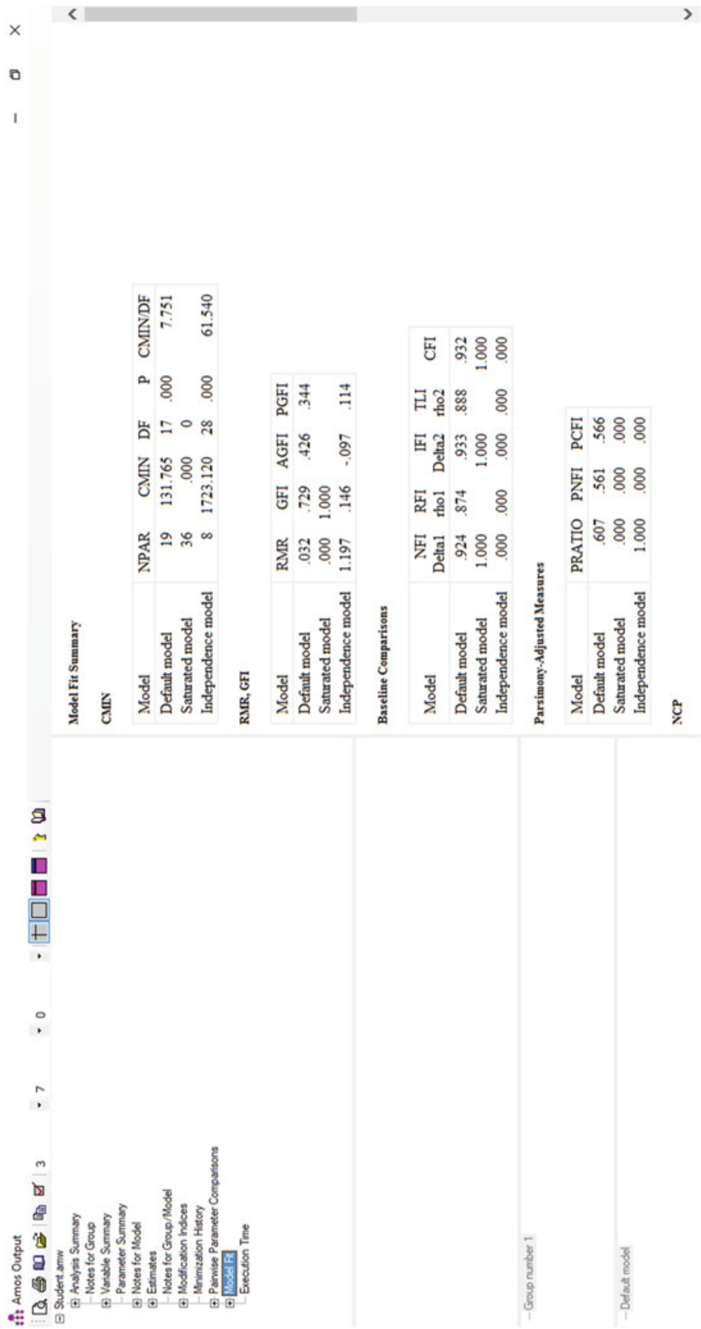
After entering the data and constructing the model, the loading factors for all the error variables are identified as 1 by default. Also, one of the loading factors between the observed variables and the latent variables is identified as 1 by default. This is done to make the model identifiable. To solve the problem, first the type of output has to be decided. This is done by checking the boxes of the ***analysis properties*** available under the ***view*** tab. Check the vacant boxes as shown in the figure. To solve the model, click the ***abacus*** icon. To see the results, click on the icon (marked in blue). In order to draw conclusions, the model has to be fit first. In the ***result output*** sheet, click on the ***model fit*** from the tree available on the left-hand side of the screen. The commonly used indices for model fit are *goodness of fit index (GFI), chi-square, root mean square error of approximation (RMSEA), incremental fit index (IFI), normal fit index (NFI) and comparative fit index (CFI)*. Observe the results obtained by the solution as shown in Fig. 4.8. The chi-square probability value is 0 which has to be >0.05; CFI, IFI, NFI values are >0.9 and are a good fit index; RMSEA value is 0.267 but has to be <0.05; and GFI is 0.729 but has to be >0.9. These indices denote that 3 out of 6 indices denote a good fit. But to ensure that all the indices denote a good fit, the software suggests few modifications. These can be viewed by clicking on the ***modification indices*** on the output page. Add the covariances as shown in Fig. 4.10. Add one by one by selecting the variables with higher modification index (MI) in order to avoid overfitting. After several iterative modifications, the modified model looks as shown in Fig. 4.10, and it is fit (Figs. 4.7, 4.9 and 4.11).

Step 7: Understanding the result

To know the correlations or the loading factors, covariances and variances, click on the ***estimates*** available on the output screen. The estimation screen provides all types of information about the basic statistics between the variables. The output screen of the estimates appears as shown in Fig. 4.12. The results can be understood by clicking on the value to be understood. When clicked, AMOS displays a pop-up which explains about the select value using its inbuilt library as shown in Fig. 4.12. Also, the estimates can be viewed by clicking view the output path diagram button as shown in Fig. 4.13 (marked in red).

Step 8: Interpreting the estimates

The loading factor between the job satisfaction and talent development is 1.145 which means for every 1 unit increase in talent development, job satisfaction increases by 1.145. Similarly, the loading factor between the job satisfaction and the flexibility is 0.23, which means for every 1 unit increase in the flexibility, job satisfaction improves by 0.23 unit. This shows that the latent variables relation as indicated by H1 and H2 are positive and can be accepted.



Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	19	131.765	17	.000	7.751
Saturated model	36	.000	0		
Independence model	8	1723.120	28	.000	61.540

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	.032	.729	.426	.344
Saturated model	.000	1.000		
Independence model	1.197	.146	-.097	.114

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.924	.874	.933	.888	.932
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	.607	.561	.566
Saturated model	.000	.000	.000
Independence model	1.000	.000	.000

NCP

Fig. 4.8 AMOS output window—model fit

Fig. 4.9 AMOS output screen—modification indices (MI)

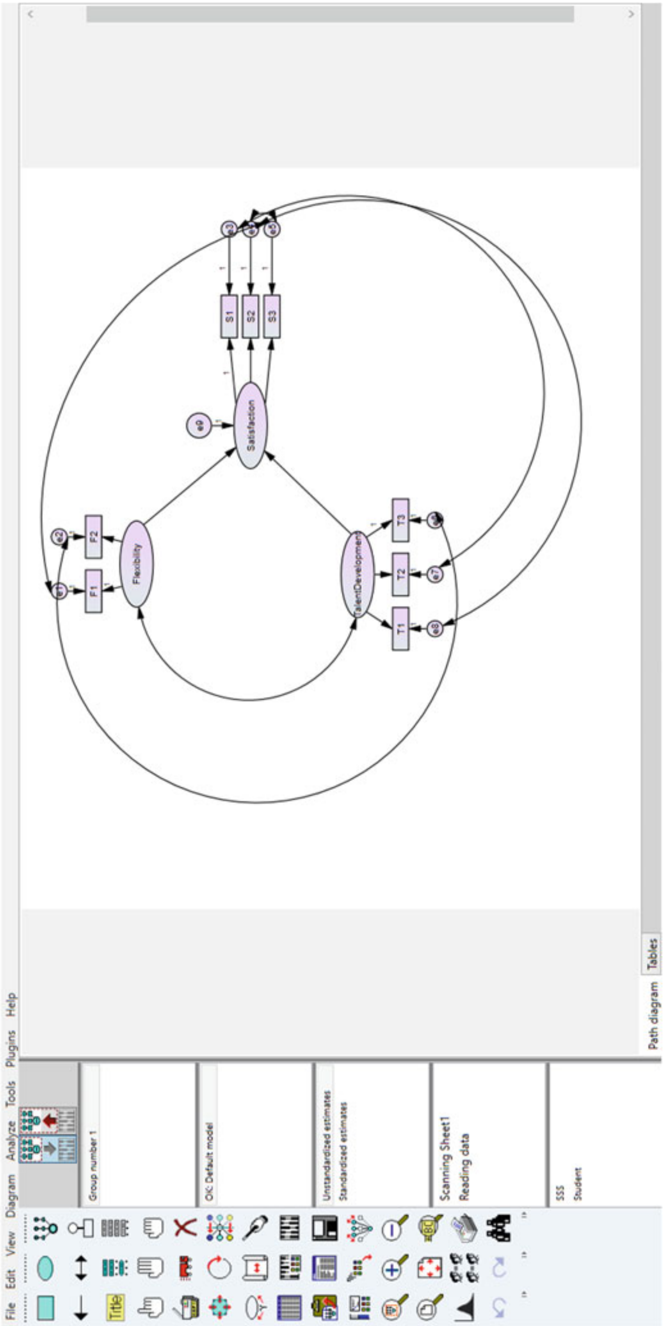
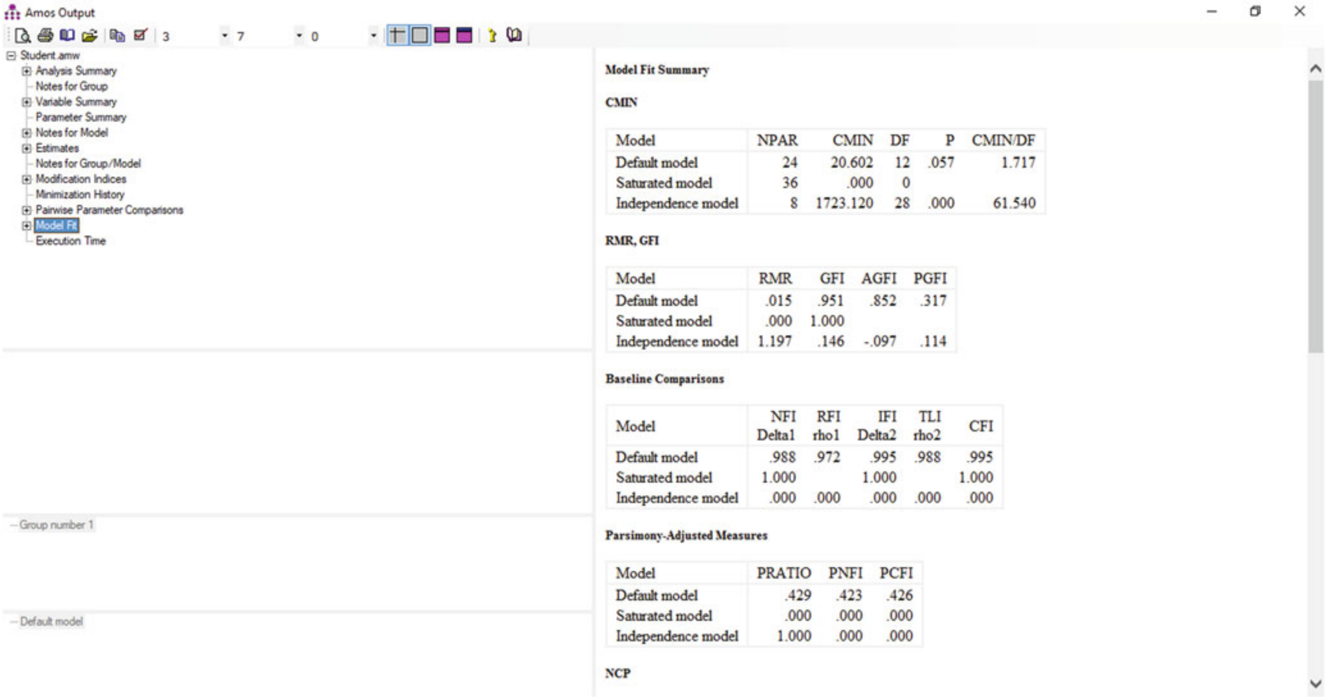


Fig. 4.10 Revised model after the modification



Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	24	20.602	12	.057	1.717
Saturated model	36	.000	0		
Independence model	8	1723.120	28	.000	61.540

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	.015	.951	.852	.317
Saturated model	.000	1.000		
Independence model	1.197	.146	-.097	.114

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.988	.972	.995	.988	.995
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	.429	.423	.426
Saturated model	.000	.000	.000
Independence model	1.000	.000	.000

NCP

Fig. 4.11 AMOS output window—model fit after the modification

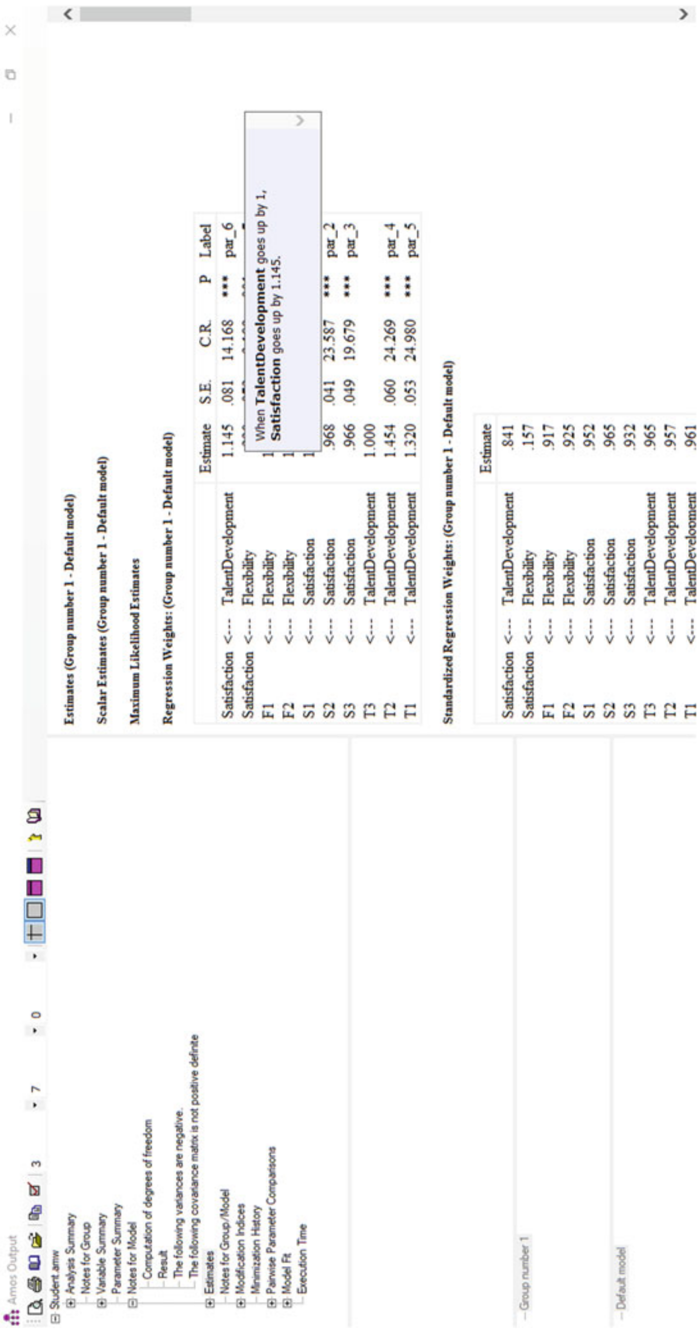


Fig. 4.12 AMOS output window—estimates

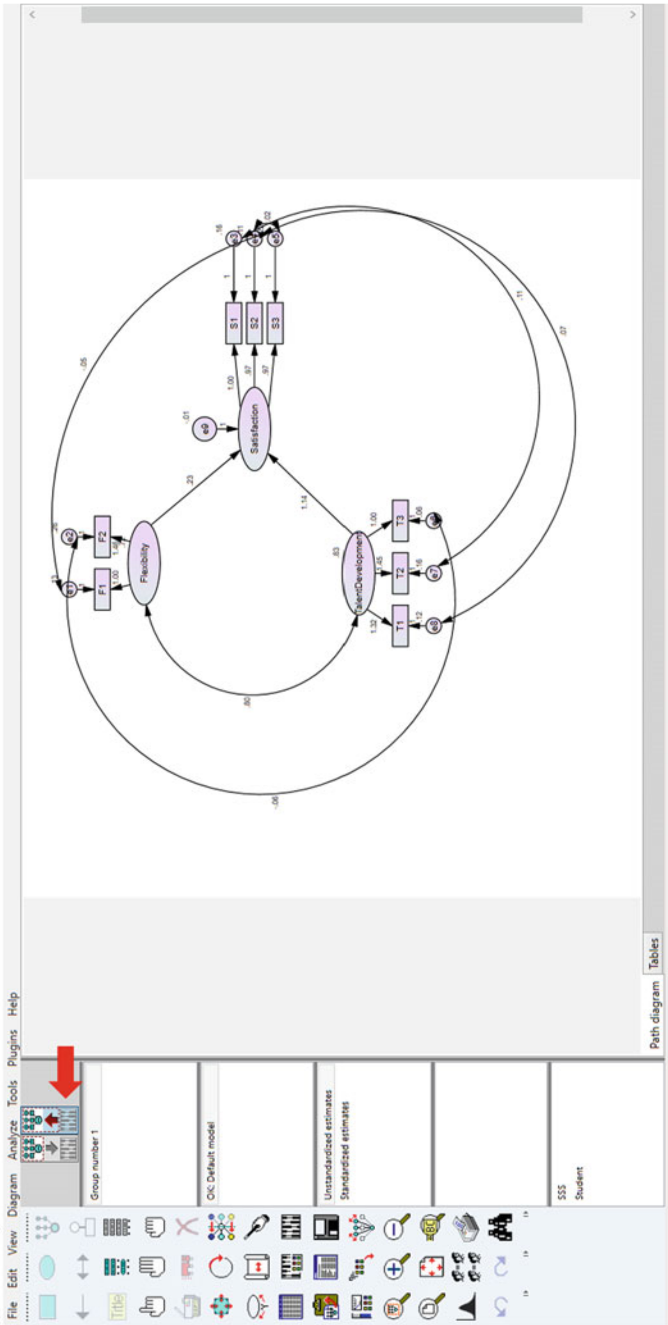


Fig. 4.13 AMOS output path diagram window

4.3 Illustrative Applications of SEM in AMOS, SPSS

4.3.1 Application 1: Healthcare System

Problem Description

As the developing technology is engulfing the human life, the human beings' life is being more prone to the technological waste. Hence, a diminishing effect of the health has been in rise from the past few years, and this has led to the increase in the healthcare service units which are hospitals, diagnostics centres, pharmacy and special care centres. Among the healthcare service units, hospitals are apparently the centre for both input and output of patients' demand.

As the business competition in this unit has been increasing, the hospitals are advancing their workforce and infrastructure towards a patient-centric healthcare services. With regard to this, the patients' willingness to visit a hospital for treatment is developing as an important criterion considering the business model of a hospital.

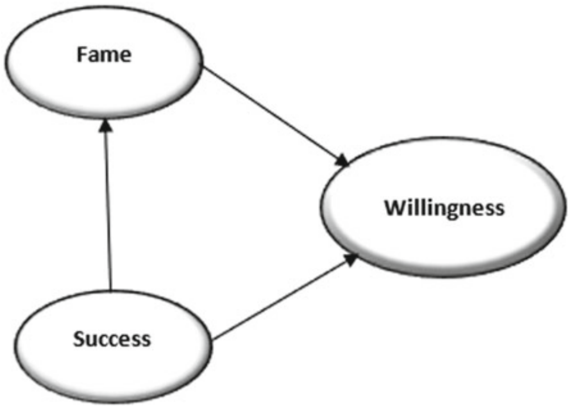
In order to analyse the willingness of the patients towards an hospital, several factors are to be considered. Usually, many variables contribute to the willingness directly and directly. As per this, hypothetical problem has been formulated considering various sources.

For a patient to visit a hospital or show his/her willingness, it is the past experience of the visit or the buzz spread around about the services in a hospital. Whenever there is no past experience, the level of fame or perception about a hospital among the people is the first line of enquiry that a patient does in order to visit that particular hospital. By consulting a group of common people and patients at a local hospital, it was confirmed that fame has a role to play in this model. But, fame is a consequence of persistent efforts from the workforce of the hospital which includes both clinical and non-clinical staff. The success in handling patients and curing them is also a factor which the patients look up to so that a hospital can be chosen for their healthcare needs. So, success of a hospital slowly builds up the fame. Therefore, success contributes to the willingness of the patient both directly and indirectly (by causing fame). This was confirmed by an enquiry study in a local hospital with a group of clinical staff, administrative staff and a few non-clinical staff. Hence, a hypothesis of relation between the willingness, fame and the success were formed and explained further.

Figure 4.14 depicts the representation of the conceptual model and interrelations of the willingness along with its contributing factors, viz. fame of the hospital and the success level of the same.

In this model, success has a direct correlation with the willingness as well as indirect through fame. The fame has a direct effect on the willingness. The problem

Fig. 4.14 Conceptual model for healthcare system



is to analyse whether this conceptual/theoretical model is accurate or not and to modify the system so that the theoretical model is apt for drawing conclusions.

Structural Model

Since the willingness, fame and success cannot be measured directly, they are labelled as latent variables. Therefore, these variables can be identified or understood using suitable identifiers, also called measured variables. According to the figure, the success and fame of the hospital influence the willingness of the patient. It can also be observed that success influences the level of perception or the fame of the hospital as well.

The identifiers or the measured variables for the willingness, fame and success are listed in Table 4.1.

The data for the identifiers are collected using a questionnaire with a rating scale from 1 to 5 where a score of 1 indicates "least considered" and a score of 5 indicates "most considered". The data collected consist of 65 values or sample size of 65. Since the willingness is caused by the variables success and fame, also, these variables are

Table 4.1 Latent variables for healthcare system

S. No.	Latent variables	Identifiers	
1	Willingness	W1	Distance
		W2	Cost for service
		W3	Ease of procedures
2	Success	S1	Facilities available
		S2	Specialties
		S3	Number of patients handled
3	Fame	F1	Leadership person
		F2	Infrastructure
		F3	Publicity

identified using many identifiers, the structural model of this problem is depicted in Fig. 4.15.

According to the procedure of structural model or drawing a path, the latent variables are mentioned in an oval. The identifiers are represented using a rectangle. The error terms are represented using a circle. The single-headed arrow indicates correlation of the causal and the caused variable. The structural model alone does not give all the info as the measured variables do have a contribution from the errors. Hence, the complete model is shown in Fig. 4.16.

The above model represents the factors responsible for errors as well.

Fig. 4.15 Structural model of healthcare system

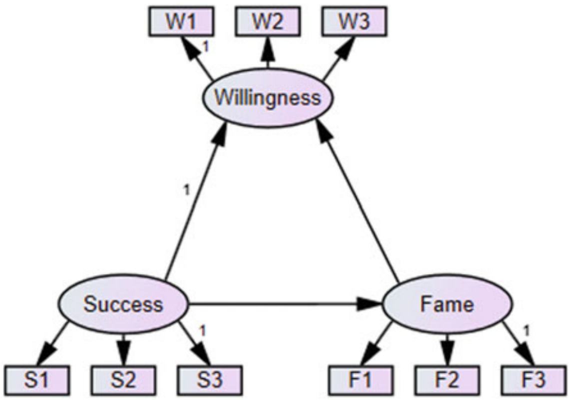
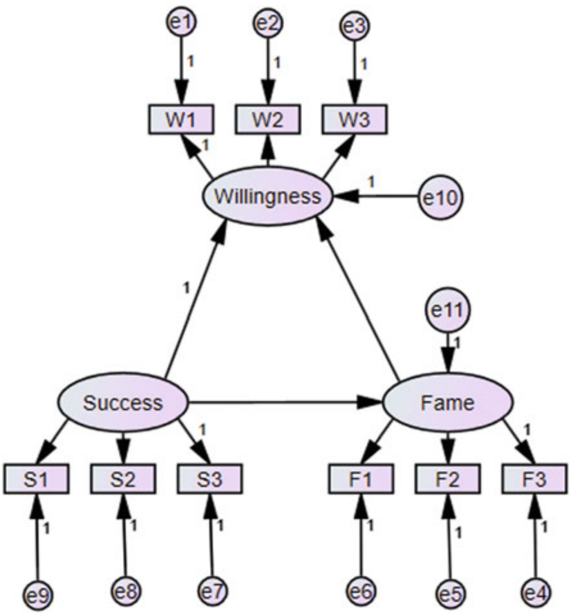


Fig. 4.16 Complete model of healthcare system



The assumptions made in this problem are:

1. The data set obtained has a normal distribution.
2. The data set has no missing data.
3. The problem has multivariate normal data.

Analysis

The theoretical model has to be analysed for its appropriateness. There are five steps to analyse the structural equation model developed. They are:

1. Model specification
2. Model identification
3. Parameter estimation
4. Assessment of model fit
5. Modification of the model.

The software used for this analysis is IBM SPSS AMOS v25.

1. ***Model specification***:

The considered model is represented previously. The variables in the model and their types are specified explicitly here. The variables—success and fame—are exogenous variables which are independent. The variables—willingness and fame—are endogenous which are dependent on other variables.

2. ***Model identification***:

This is done to check whether the unknown parameters can be estimated using the number of knowns or not. If "s" is the number of measured variables, the number of known parameters is $½(s + 1) * s$. In the model shown above, the number of known parameters is 37. The parameters to be estimated are 3 latent variable variance, 11 error variance and 23 loading factors. This accounts to 37 unknown parameters. This is a case of just-identified model. But, the degree of freedom $37 - 37 = 0$. Hence, few values are imputed to 1 to gain few degrees of freedom. Therefore, the unknown parameters can be estimated.

3. ***Estimation of parameters, model fit and modification***:

The parameters are estimated by using a covariance matrix by the software programme. It is estimated for a fit model only. The initial assessment revealed the following output in Fig. 4.17.

The chi-square value of 37.5 with 25 degrees of freedom had a p value of under 0.05 which is less than the threshold 0.05. The root mean square error of approximation (RMSEA) has a value 0.088 which is more than the prescribed threshold 0.05. The goodness of fit index (GFI) has a value of 0.898 which had to be more than 0.9. So, these values indicate that the model is not fit.

But, the software programme suggests a modification index which can increase the model fitness. The modification index indicates to have a covariance between e1&e5 and e6&e9.

Model	RMR	GFI	AGFI	PGFI
Default model	.051	.898	.816	.499
Saturated model	.000	1.000		
Independence model	.056	.867	.833	.693

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.088	.000	.144	.148
Independence model	.077	.000	.125	.203

Fig. 4.17 Estimated parameters for healthcare system

The estimation, assessing model fitness and modification are iterative in AMOS. After adding a covariance relation between e1&e5 and e6&e9, the output of the model fitness analysis after the change is shown in Fig. 4.18.

After the modification, chi-square value of 22.3 with degree of freedom 23 has a p value of 0.5 which is greater than 0.05. The GFI has a value 0.928 which is greater than 0.9, and RMSEA has a value 0 which is less than 0.05. All these values indicate that a model is fit.

After this modification, the model looks as shown in Fig. 4.19.

The estimates of the unknown parameters:
The regression weights or the load factors are shown in Fig. 4.20.
Estimates for covariances (Fig. 4.21).
Estimates for the variances (Fig. 4.22).
Estimates for residual covariances (Fig. 4.23).

Assessing the model fitness with fitness indices:
There are two types of indices to check the model fit. They are absolute indices and relative indices. The indices calculated by the software are shown in Table 4.2.

Model	RMR	GFI	AGFI	PGFI
Default model	.043	.928	.859	.474
Saturated model	.000	1.000		
Independence model	.056	.867	.833	.693

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.000	.000	.099	.689
Independence model	.077	.000	.125	.203

Fig. 4.18 Model fitness analysis for healthcare system

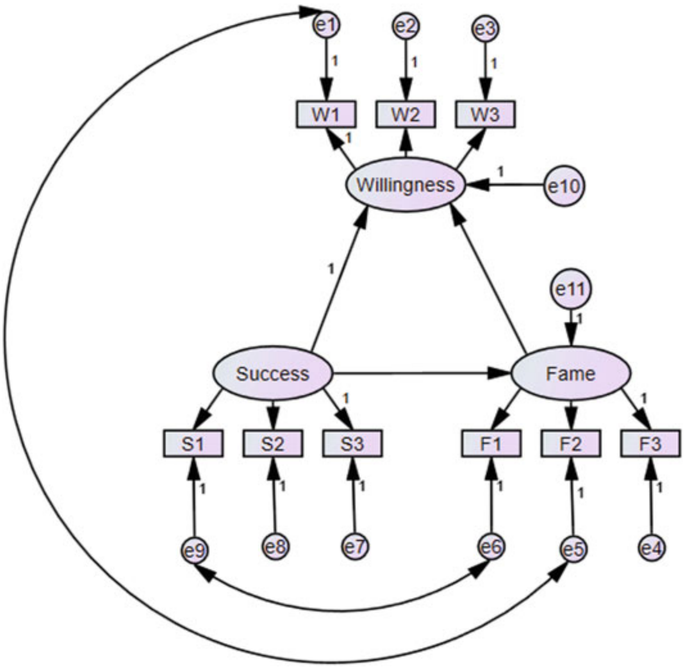


Fig. 4.19 Modified structural model of healthcare system

			Estimate	S.E.	C.R.	P
Fame	<---	Success	-.269	.257	-1.044	.297
Willingness	<---	Success	1.000			
Willingness	<---	Fame	2.241	2.065	1.085	.278
W1	<---	Willingness	1.000			
W2	<---	Willingness	-.943	1.021	-.923	.356
W3	<---	Willingness	-2.986	2.841	-1.051	.293
F3	<---	Fame	1.000			
F2	<---	Fame	-3.087	2.511	-1.230	.219
F1	<---	Fame	-.372	.984	-.378	.705
S3	<---	Success	1.000			
S2	<---	Success	.618	.630	.982	.326
S1	<---	Success	.578	.460	1.257	.209

Fig. 4.20 Regression weights for healthcare system

Fig. 4.21 Covariances for healthcare system

			Estimate	S.E.	C.R.	P	Label
e1	<-->	e5	.155	.061	2.568	.010	
e6	<-->	e9	-.116	.059	-1.959	.050	

	Estimate	S.E.	C.R.	P	Label
Success	.074	.081	.918	.358	
e11	.010	.013	.809	.418	
e10	-.053	.064	-.831	.406	
e1	.284	.053	5.367	***	
e2	.236	.043	5.479	***	
e3	.649	.159	4.072	***	
e4	.230	.042	5.451	***	
e5	.220	.130	1.693	.090	
e6	.665	.118	5.649	***	
e7	.636	.130	4.882	***	
e8	.713	.132	5.419	***	
e9	.297	.059	5.071	***	

Fig. 4.22 Estimates of variances for healthcare system

	S1	S2	S3	F1	F2	F3	W3	W2	W1
S1	-.001								
S2	.004	.000							
S3	.028	-.097	.000						
F1	-.001	-.064	.084	.002					
F2	-.034	-.004	.017	-.057	-.012				
F3	.010	-.085	-.025	-.064	-.002	.000			
W3	-.024	.045	-.021	-.101	-.010	.045	.000		
W2	-.016	-.065	.075	.048	-.001	-.002	-.027	.000	
W1	.026	-.015	-.056	-.105	-.012	-.006	-.010	-.013	-.001

Fig. 4.23 Regression residual covariances for healthcare system

Table 4.2 Fit indices for healthcare system

S. No.	Type	Index	Actual value	Preferred value	Conclusion
1	Absolute	Chi-square (χ^2)	0.5	>0.05	Good fit
		Root mean square error of approximation (RMSEA)	0	<0.05	Good fit
		Goodness of fit (GFI)	0.928	>0.9	Good fit
2	Relative	Normal fit (NFI)	0.549	>0.9	–
		Incremental fit (IFI)	1	>0.9	Good fit
		Comparative fit (CFI)	1.025	>0.9	Good fit

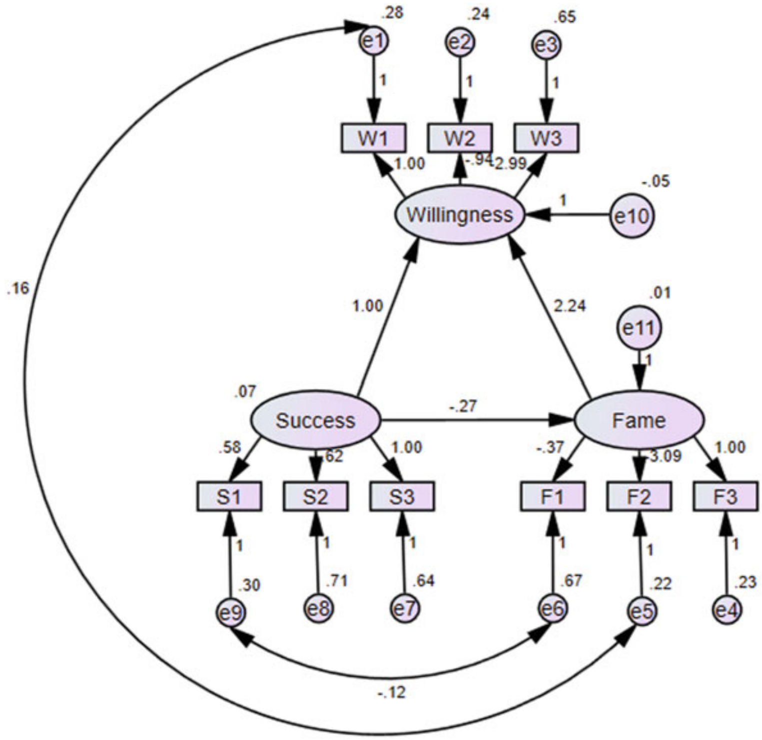


Fig. 4.24 SEM model with estimated parameters for healthcare system

The model with all the estimates will be represented as shown in Fig. 4.24.

Results and Conclusions
As per the initial model, the notion of the modeller was that willingness had a positive effect by success and fame. Also, success influenced fame. After the result, loading factors indicate that the notion of the modeller of success and fame strongly influences with a value of 1 and 2.24, respectively. But, the success has a negative effect on the fame as per the data which are quite surprising. It is also seen that the fame and infrastructure has a negative relation with the leadership which is quite reasonable from the responders' view. There is a negative effect between the procedure and cost as well which makes the distance as a positive effect. Since the *p* value of the estimates (non-standardized) is all greater than significance level 0.05 for the factor loadings, the estimates do account for large variance and hence cannot be supported.

4.3.2 Application 2: Marketing Model

Problem Description

A product or a service is essential for a person to carry out an activity which is not possible to complete by bare-handed. For this purpose, many industries are understanding the needs of the people and churning out products which are being helpful for the people to carry out their chores and do the tasks effectively. As and when the industries seek opportunities in these products to magnify their bottom line, more and more products rise every day. This leads to effective competition, and hence, the necessity for the marketing has rose. The main purpose is to reach the customer. Figure 4.25 represents conceptual model and interrelations of the customer reach with its related factors, viz. mode of publicity, social influence and competition.

In this model, mode of publicity, social influence and competition have a direct influence on the customer reach. The problem is to analyse whether this conceptual/theoretical model is accurate or not and to modify the system so that the theoretical model is apt for drawing conclusions. Therefore, three hypotheses have been postulated.

H1: Mode of publicity has a positive influence for a better customer reach.
H2: Social influence can increase the customer reach.
H3: More competition can decrease the customer reach.

Structural Model

Since the customer reach, social influence, mode of publicity and competition cannot be measured directly, they are labelled as latent variables. Therefore, these variables can be identified or understood using suitable identifiers, also called measured variables. The identifiers or the measured variables for the customer reach, mode of publicity, social influence and competition are listed in Table 4.3.

Fig. 4.25 Conceptual model for marketing system

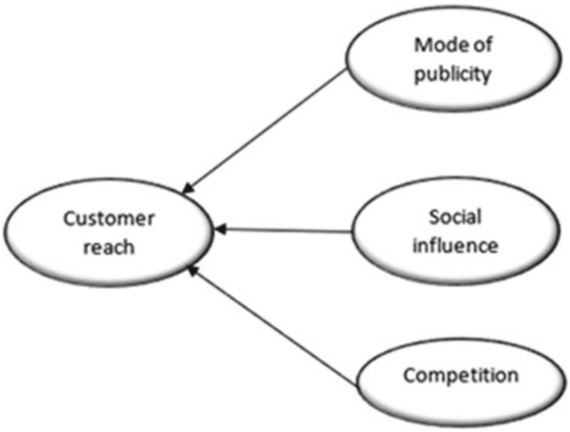


Table 4.3 Latent variables for marketing system

S. No.	Latent variables	Identifiers	
1	Customer reach	R1	Budget for publicity
		R2	Impact of Idea of product
		R3	Human resource for marketing
2	Mode of publicity	M1	Print media usage
		M2	TV usage
		M3	Online platforms usage
3	Social influence	S1	Others buying the product
		S2	Prestige of acquiring
		S3	Necessity of a product
4	Competition	C1	Number of similar products
		C2	Cost of competing products
		C3	Quality of competing products

The data for the identifiers are collected using a questionnaire with a rating scale from 1 to 5 where a score of 1 indicates "strongly disagree" and a score of 5 indicates "strongly agree". The data collected consist of 99 values or sample size of 99. The structural model of this problem is depicted in Fig. 4.26.

According to the procedure of structural model or drawing a path, the latent variables are mentioned in an oval. The identifiers are represented using a rectangle. The error terms are represented using a circle. The single-headed arrow indicates correlation of the causal and the caused variable.

The structural model alone does not give all the info as the measured variables do have a contribution from the errors. Hence, the complete model is shown in Fig. 4.27.

The above model represents the factors responsible for errors as well.

The assumptions made in this problem are:

4. The data set obtained has a normal distribution.
5. The problem has multivariate normal data.

Analysis

The theoretical model has to be analysed for its appropriateness. There are five steps to analyse the structural equation model developed. They are:

6. Model specification
7. Model identification
8. Parameter estimation

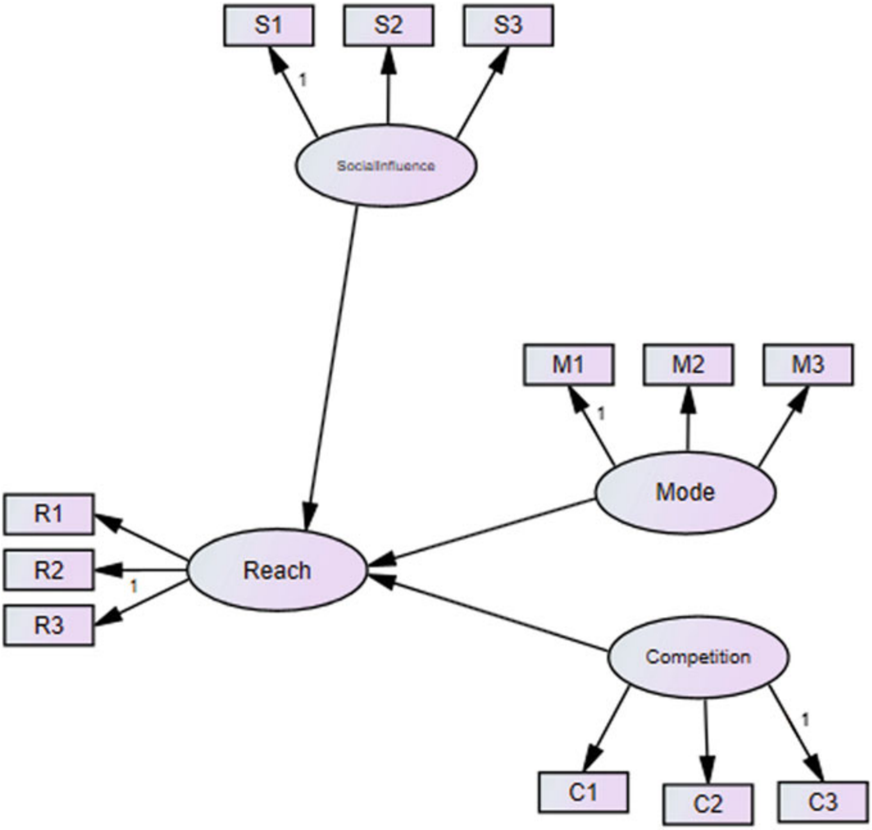


Fig. 4.26 Structural model of marketing system

9. Assessment of model fit
10. Modification of the model.

The software used for this analysis is IBM SPSS AMOS v25.

4. ***Model specification***:

The considered model is represented previously. The variables in the model and their types are specified explicitly here. The variables—mode of publicity, competition and social influence—are exogenous variables which are independent. The variable—customer reach—is endogenous which is dependent on other variables.

5. ***Model identification***:

This is done to check whether the unknown parameters can be estimated using the number of knowns or not. If "s" is the number of measured variables, the number of known parameters is ½($s + 1$) * s. In the model shown above, the number of known

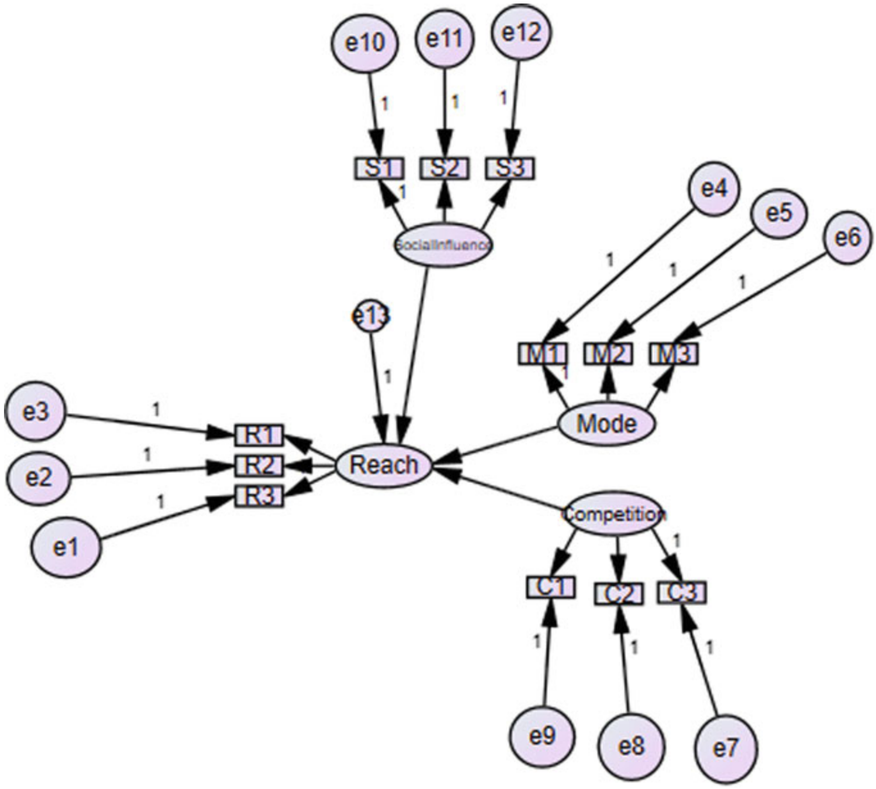


Fig. 4.27 Complete model of marketing system

parameters is 78. The parameters to be estimated are 4 latent variable variance, 13 error variance and 11 loading factors. This accounts to 38 unknown parameters. The degree of freedom $78 - 38 = 40$.

6. ***Estimation of parameters, model fit and modification***:

The parameters are estimated by using a covariance matrix by the software programme. It is estimated for a fit model only. The initial assessment revealed the following output in Fig. 4.28.

The chi-square value of 100 with 74 degrees of freedom had a p value of under 0.05 which is less than the threshold 0.05. The root mean square error of approximation (RMSEA) has a value 0.06 which is more than the prescribed threshold 0.05. The goodness of fit index (GFI) has a value of 0.877 which had to be more than 0.9. So, these values indicate that the model is not fit. But, the software programme suggests a modification index which can increase the model fitness. The estimation, assessing model fitness and modification are iterative in AMOS.

After adding a covariance relation between suggested variables, the output of the model fitness analysis after the change is shown in Fig. 4.29.

Fig. 4.28 Estimated parameters for marketing system

Model	RMR	GFI	AGFI	PGFI
Default model	.170	.877	.871	.832
Saturated model	.000	1.000		
Independence model	.203	.821	.789	.695

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.264	.344	.580	.668	.628
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.060	.023	.088	.285
Independence model	.104	.079	.129	.001

Fig. 4.29 Model fitness analysis for marketing system

Model	RMR	GFI	AGFI	PGFI
Default model	.119	.941	.882	.470
Saturated model	.000	1.000		
Independence model	.203	.821	.789	.695

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.694	.482	.973	.937	.963
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.026	.000	.076	.728
Independence model	.104	.079	.129	.001

After the modification, chi-square value of 41.584 with degree of freedom 39 has a p value of 0.369 which is greater than 0.05. The GFI has a value 0.941 which is greater than 0.9, and RMSEA has a value 0.026 which is less than 0.05. All these values indicate that a model is fit. After this modification, the model looks as shown in Fig. 4.30.

The estimates of the unknown parameters:
The regression weights or the load factors are calculated in Fig. 4.31.
Estimates for covariances (Fig. 4.32).
Estimates for the variances (Fig. 4.33).
Estimates for total effects (Fig. 4.34).

Assessing the model fitness with fitness indices:
There are two types of indices to check the model fit. They are absolute indices and relative indices. The indices calculated by the software are shown in Table 4.4.

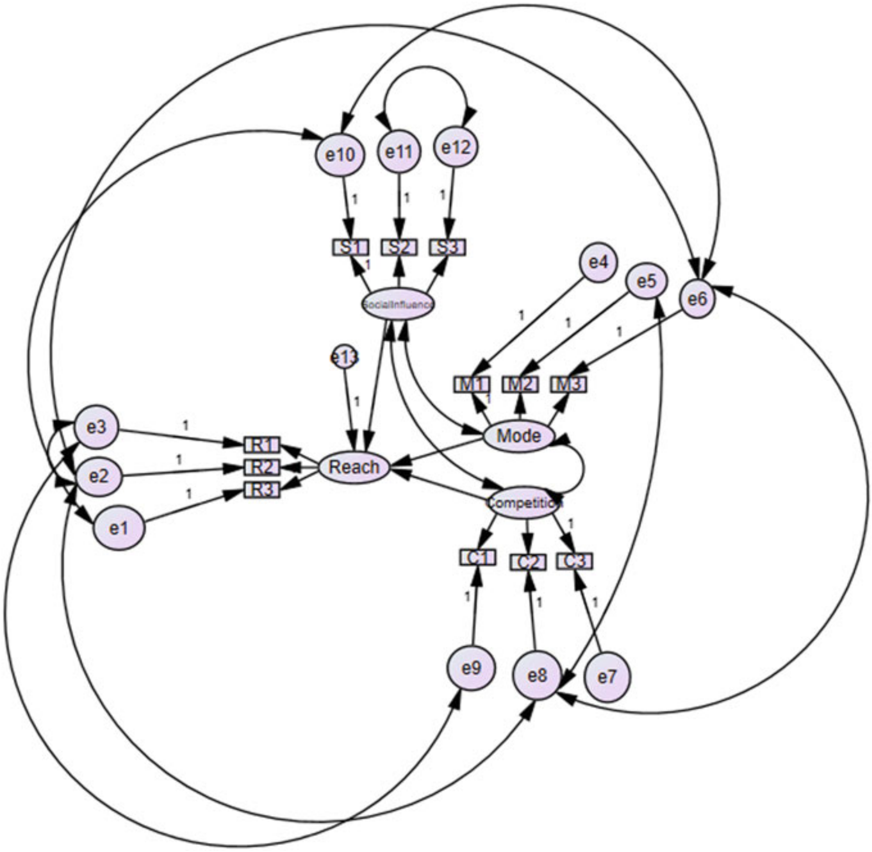


Fig. 4.30 Modified structural model of marketing system

The model with all the estimates will be represented as shown in Fig. 4.35.

Results and Conclusions

As per the initial model, the hypothesis of the modeller was that social influence and mode of publicity positively influence customer reach and the competition negatively influences the customer reach. After the result, loading factors indicate that the competition and social influence positively influence the customer reach and mode of publicity negatively influences the customer reach. Since the *p* value of the estimates (non-standardized) is all greater than significance level 0.05 for the factor loadings, the estimates do account for large variance and hence cannot be supported. Finally, the hypothesis results are as follows:

H1: Reject
H2: Accept
H3: Reject.

			Estimate	S.E.	C.R.	P	Label
Reach	<---	Mode	-1.037	.733	-1.413	.158	par_9
Reach	<---	Competition	.086	.185	.465	.642	par_10
Reach	<---	SocialInfluence	.060	.433	.138	.890	par_11
R3	<---	Reach	1.000				
R2	<---	Reach	-1.429	.840	-1.700	.089	par_1
R1	<---	Reach	-.972	.745	-1.304	.192	par_2
S1	<---	SocialInfluence	1.000				
S2	<---	SocialInfluence	-.280	.738	-.380	.704	par_3
S3	<---	SocialInfluence	-2.051	1.494	-1.373	.170	par_4
M1	<---	Mode	1.000				
M2	<---	Mode	3.847	2.702	1.424	.155	par_5
M3	<---	Mode	1.081	.758	1.426	.154	par_6
C3	<---	Competition	1.000				
C2	<---	Competition	.580	.488	1.190	.234	par_7
C1	<---	Competition	-.283	.268	-1.056	.291	par_8

Fig. 4.31 Regression weights for marketing system

			Estimate	S.E.	C.R.	P	Label
SocialInfluence	<-->	Mode	.035	.038	.941	.347	par_12
Mode	<-->	Competition	.100	.080	1.247	.213	par_13
SocialInfluence	<-->	Competition	-.147	.114	-1.292	.196	par_19
e5	<-->	e8	-.516	.269	-1.923	.055	par_14
e3	<-->	e9	-.495	.176	-2.821	.005	par_15
e2	<-->	e3	.370	.147	2.513	.012	par_16
e11	<-->	e12	.376	.181	2.074	.038	par_17
e1	<-->	e10	.369	.155	2.371	.018	par_18
e2	<-->	e8	-.355	.160	-2.218	.027	par_20
e6	<-->	e10	-.339	.150	-2.256	.024	par_21
e2	<-->	e6	.305	.120	2.548	.011	par_22
e6	<-->	e8	.203	.185	1.100	.271	par_23

Fig. 4.32 Covariances for marketing system

4.3.3 Application 3: Productivity Model

Problem Description

It has become a usual practice in almost all the sectors of profession to improve the productivity. In this line, productivity can be extended to the entities like labour, machine. Productivity can be defined by the ratio of output to the input and multiplied by 100 whenever the expectation is in terms of percentage. The current problem deals

	Estimate	S.E.	C.R.	P	Label
SocialInfluence	.037	.088	.423	.673	par_24
Mode	.072	.078	.928	.353	par_25
Competition	-.165	.319	-.518	.605	par_26
e13	-.025	.080	-.309	.757	par_27
e1	1.373	.206	6.655	***	par_28
e2	.831	.179	4.650	***	par_29
e3	1.824	.265	6.880	***	par_30
e4	1.457	.212	6.869	***	par_31
e5	.465	.645	.721	.471	par_32
e6	1.501	.222	6.768	***	par_33
e7	2.252	.469	4.799	***	par_34
e8	1.887	.294	6.417	***	par_35
e9	1.731	.250	6.914	***	par_36
e10	1.598	.240	6.663	***	par_37
e11	1.841	.263	6.998	***	par_38
e12	1.577	.383	4.120	***	par_39

Fig. 4.33 Estimates of variances for marketing system

Fig. 4.34 Estimates of total effects for marketing system

	Competition	Mode	SocialInfluence	Reach
Reach	.086	-1.037	.060	.000
C1	-.283	.000	.000	.000
C2	.580	.000	.000	.000
C3	1.000	.000	.000	.000
M3	.000	1.081	.000	.000
M2	.000	3.847	.000	.000
M1	.000	1.000	.000	.000
S3	.000	.000	-2.051	.000
S2	.000	.000	-.280	.000
S1	.000	.000	1.000	.000
R1	-.083	1.007	-.058	-.972
R2	-.123	1.481	-.085	-1.429
R3	.086	-1.037	.060	1.000

with a model involving labour productivity. It can be observed that labour productivity depends on the environment of working and personal satisfaction of the labourers. Also, environment might influence the personal satisfaction to some extent.

Figure 4.36 depicts the representation of the conceptual model and interrelations of the labour productivity along with its contributing factors, viz. environment in which the labourer works and personal satisfaction of the labourer.

Table 4.4 Fit indices for marketing system

S. No.	Type	Index	Actual value	Preferred value	Conclusion
1	Absolute	Chi-square (χ^2)	0.359	>0.05	Good fit
		Root mean square error of approximation (RMSEA)	0	<0.05	Good fit
		Goodness of fit (GFI)	0.941	>0.9	Good fit
2	Relative	Normal fit (NFI)	0.694	>0.9	Acceptable
		Incremental fit (IFI)	0.973	>0.9	Good fit
		Comparative fit (CFI)	0.963	>0.9	Good fit

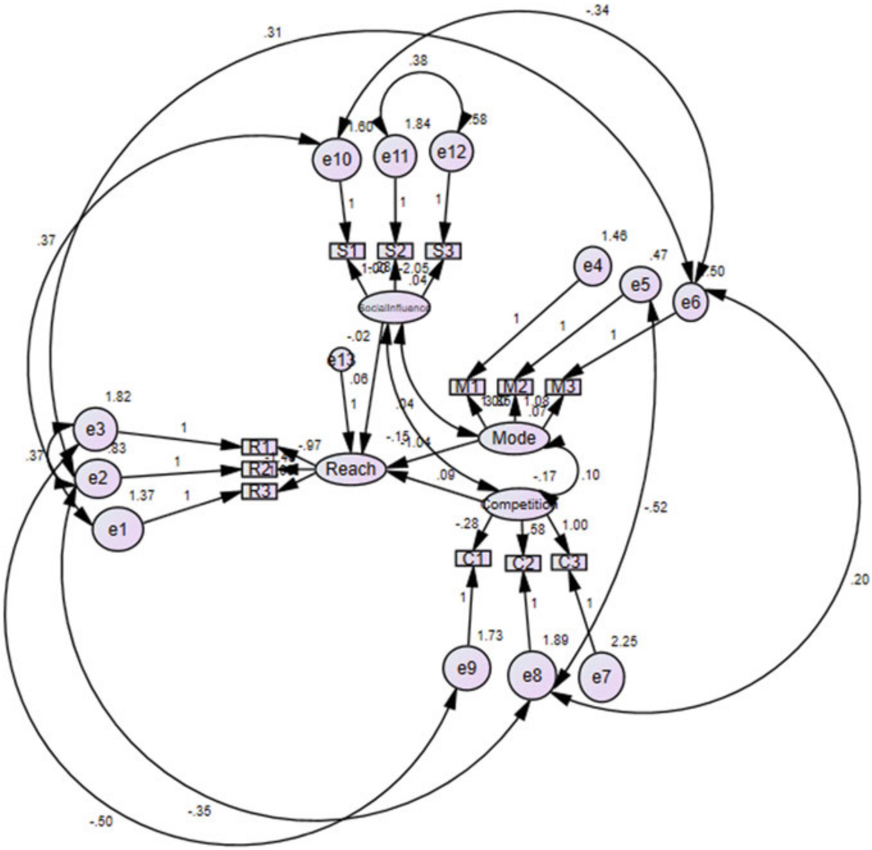


Fig. 4.35 SEM model with estimated parameters for marketing system

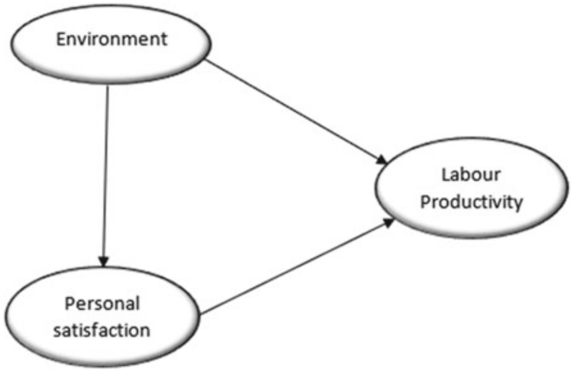


Fig. 4.36 Conceptual model for productivity analysis

In this model, environment has a direct correlation with the labour productivity as well as indirect correlation through personal satisfaction. The personal satisfaction has a direct effect on the labour productivity. The problem is to analyse whether this conceptual/theoretical model is accurate or not and to modify the system so that the theoretical model is apt for drawing conclusions.

Therefore, the following hypothesis is postulated.

H1: Good environment has a positive influence on the productivity.
H2: Good environment has a positive influence on the satisfaction.
H3: Personal satisfaction has a positive influence on the productivity.

Structural Model

Since the labour productivity, environmental condition and personal satisfaction cannot be measured directly, they are labelled as latent variables. Therefore, these variables can be identified or understood using suitable identifiers, also called measured variables. The identifiers or the measured variables for the labour productivity, environment and personal satisfaction are given in Table 4.5.

The data for the identifiers are collected using a questionnaire with a rating scale from 1 to 5 where a score of 1 indicates "disagree" and a score of 5 indicates "agree". The data collected consist of 113 values or sample size of 113. Since the labour productivity is caused by the variables environment and personal satisfaction, also, these variables are identified using many identifiers, the structural model of this problem is depicted in Fig. 4.37.

According to the procedure of structural model or drawing a path, the latent variables are mentioned in an oval. The identifiers are represented using a rectangle. The error terms are represented using a circle. The single-headed arrow indicates correlation of the causal and the caused variable. The structural model alone does not give all the info as the measured variables do have a contribution from the errors. Hence, the complete model is shown in Fig. 4.38.

The above model represents the factors responsible for errors as well.

Table 4.5 Latent variables for productivity analysis

S. No.	Latent variables	Identifiers	
1	Labour productivity	L1	Quantity of input consumption
		L2	Sales generated
		L3	Discipline among the workers
		L4	Number of tasks done
		L5	Interest developed in the work
2	Environment	E1	Lighting conditions in the workplace
		E2	Noise
		E3	Safety measures taken
		E4	Co-workers relationship
3	Personal satisfaction	P1	Commuting time spent
		P2	Health condition
		P3	Family commitments

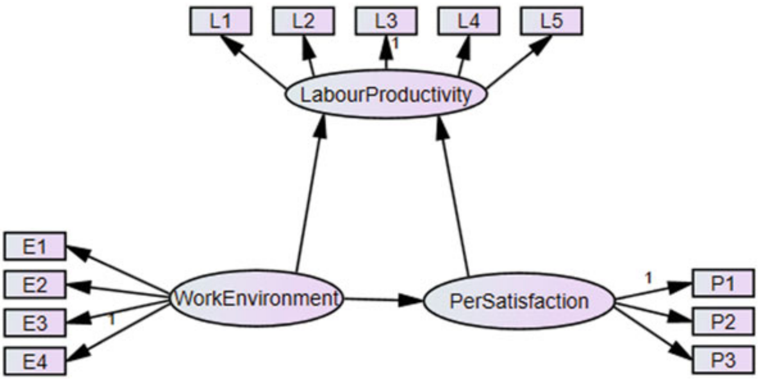


Fig. 4.37 Structural model for productivity analysis

The assumptions made in structural equation modelling problem are:

6. The data set obtained has a normal distribution.
7. The data set has no missing data.
8. The problem has multivariate normal data.

Analysis

The theoretical model has to be analysed for its appropriateness. There are five steps to analyse the structural equation model developed. They are:

11. Model specification
12. Model identification
13. Parameter estimation

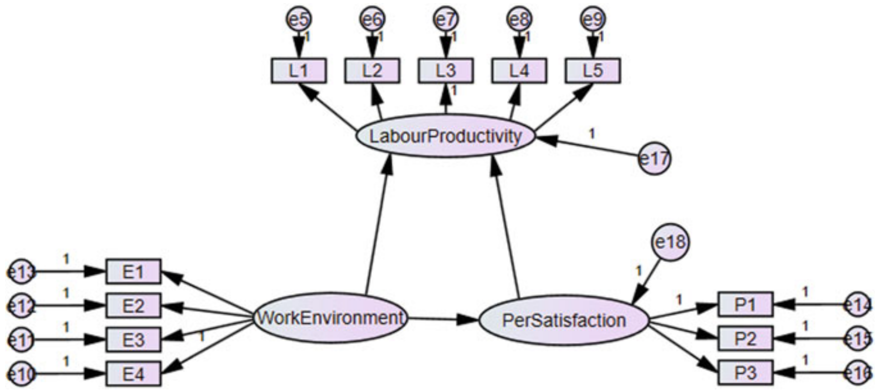


Fig. 4.38 Complete model of productivity analysis

14. Assessment of model fit
15. Modification of the model.

The software used for this analysis is IBM SPSS AMOS v25.

7. ***Model specification***:

The considered model is represented previously. The variables in the model and their types are specified explicitly here. The variable—work environment—is an exogenous variable which is independent. The variables—labour productivity and satisfaction—are endogenous which are dependent on other variables.

8. ***Model identification***:

This is done to check whether the unknown parameters can be estimated using the number of knowns or not. If "s" is the number of measured variables, the number of known parameters is ½($s + 1$) * s. In the model shown above, the number of known parameters is 78. The parameters to be estimated are 3 latent variable variance, 14 error variance and 12 loading factors. This accounts to 29 unknown parameters.

9. ***Estimation of parameters, model fit and modification***:

The parameters are estimated by using a covariance matrix by the software programme. It is estimated for a fit model only. The initial assessment revealed the following output Fig. 4.39.

The chi-square value of 106.1 with 78 degrees of freedom had a p value of under 0.05 which is less than the threshold 0.05. The root mean square error of approximation (RMSEA) has a value 0.056 which is more than the prescribed threshold 0.05. The goodness of fit index (GFI) has a value of 0.87 which had to be more than 0.9. So, these values indicate that the model is not fit.

But, the software programme suggests a modification index which can increase the model fitness.

Model	RMR	GFI	AGFI	PGFI
Default model	.184	.870	.870	.870
Saturated model	.000	1.000		
Independence model	.236	.749	.703	.634

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.547	.617	.820	.859	.833
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.056	.024	.081	.348
Independence model	.148	.128	.169	.000

Fig. 4.39 Estimated parameters for productivity analysis

The estimation, assessing model fitness and modification are iterative in AMOS.

After adding a covariance relation between suggested variables, the output of the model fitness analysis after the change is shown in Fig. 4.40.

After the modification, chi-square value of 41.7 with degree of freedom 44 has a p value of 0.569 which is greater than 0.05. The GFI has a value 0.946 which is greater than 0.9, and RMSEA has a value 0 which is less than 0.05. All these values indicate that a model is fit.

After this modification, the model looks as shown in Fig. 4.41.

Model	RMR	GFI	AGFI	PGFI
Default model	.105	.946	.904	.533
Saturated model	.000	1.000		
Independence model	.236	.749	.703	.634

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.822	.733	1.012	1.020	1.000
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.000	.000	.057	.909
Independence model	.148	.128	.169	.000

Fig. 4.40 Model fitness for productivity analysis

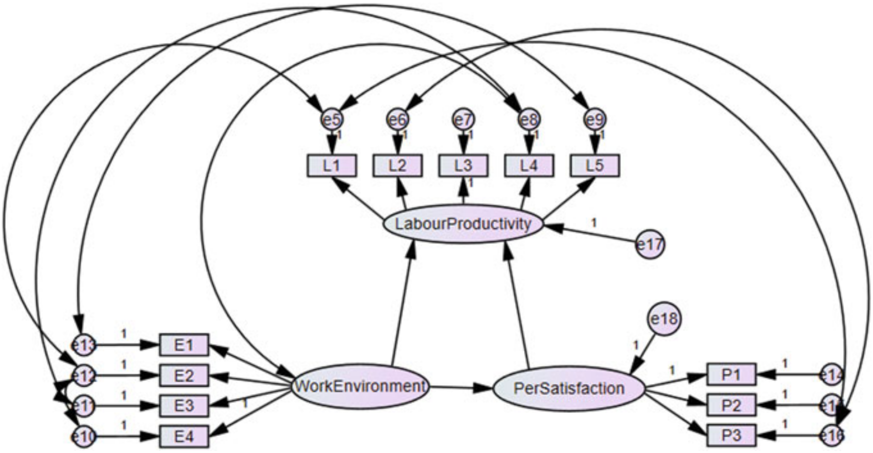


Fig. 4.41 Modified structural model for productivity analysis

The estimates of the unknown parameters:
The regression weights or the load factors are calculated in Fig. 4.42.
Estimates for covariances (Fig. 4.43).
Estimates for the variances (Fig. 4.44).
Estimates for total effects (Fig. 4.45).

			Estimate	S.E.	C.R.	P	Label
PerSatisfaction	<---	Environment	6.141	5.373	1.143	.253	
LabourProductivity	<---	Environment	-1.046	2.485	-.421	.674	
LabourProductivity	<---	PerSatisfaction	.239	.277	.862	.389	
L1	<---	LabourProductivity	.727	.557	1.304	.192	
L2	<---	LabourProductivity	-1.308	.713	-1.834	.067	
L3	<---	LabourProductivity	1.000				
L4	<---	LabourProductivity	2.344	2.424	.967	.334	
L5	<---	LabourProductivity	-.182	.772	-.235	.814	
E4	<---	Environment	1.000				
E3	<---	Environment	-.419	.712	-.588	.557	
E2	<---	Environment	-2.249	1.848	-1.217	.224	
E1	<---	Environment	-2.553	2.086	-1.224	.221	
P1	<---	PerSatisfaction	1.000				
P2	<---	PerSatisfaction	-.485	.087	-5.586	***	
P3	<---	PerSatisfaction	-.354	.065	-5.419	***	

Fig. 4.42 Regression weights for productivity analysis

			Estimate	S.E.	C.R.	P	Label
e9	<-->	e13	-.609	.243	-2.502	.012	
e8	<-->	Environment	-.116	.126	-.921	.357	
e8	<-->	e10	-.366	.173	-2.118	.034	
e6	<-->	e16	.167	.073	2.277	.023	
e5	<-->	e12	.279	.131	2.139	.032	
e5	<-->	e16	-.222	.080	-2.759	.006	
e11	<-->	e12	-.252	.125	-2.006	.045	

Fig. 4.43 Covariance for productivity analysis

	Estimate	S.E.	C.R.	P	Label
Environment	.033	.054	.623	.533	
e18	.627	.591	1.060	.289	
e17	.047	.079	.593	.553	
e5	1.477	.198	7.460	***	
e6	1.151	.183	6.276	***	
e7	1.381	.192	7.191	***	
e8	1.349	.326	4.145	***	
e9	4.294	.564	7.613	***	
e10	1.972	.260	7.583	***	
e11	1.403	.184	7.612	***	
e12	1.282	.182	7.046	***	
e13	1.480	.214	6.925	***	
e14	-.302	.233	-1.294	.196	
e15	.888	.126	7.031	***	
e16	.537	.075	7.196	***	

Fig. 4.44 Estimates for variances for productivity analysis

Assessing the model fitness with fitness indices:
There are two types of indices to check the model fit. They are absolute indices and relative indices. The values of calculated indices are given in Table 4.6.

The model with all the estimates will be represented as shown in Fig. 4.46.

Results and Conclusions
As per the initial model, the hypothesis of the modeller was that personal satisfaction and environment positively influenced the productivity. Also, personal satisfaction positively got influenced by the environment. After the result, loading factors indicate that the hypothesis of the modeller of personal satisfaction influences productivity with a value of 0.24, and environment influences strongly the satisfaction with a

	Environment	PerSatisfaction	LabourProductivity
PerSatisfaction	6.141	.000	.000
LabourProductivity	.421	.239	.000
P3	-2.174	-.354	.000
P2	-2.980	-.485	.000
P1	6.141	1.000	.000
E1	-2.553	.000	.000
E2	-2.249	.000	.000
E3	-.419	.000	.000
E4	1.000	.000	.000
L5	-.076	-.043	-.182
L4	.986	.560	2.344
L3	.421	.239	1.000
L2	-.550	-.312	-1.308
L1	.306	.174	.727

Fig. 4.45 Estimates of total effects for productivity analysis

Table 4.6 Fit indices for productivity analysis

S. No.	Type	Index	Actual value	Preferred value	Conclusion
1	Absolute	Chi-square (χ^2)	0.569	>0.05	Good fit
		Root mean square error of approximation (RMSEA)	0.000	<0.05	Good fit
		Goodness of fit (GFI)	0.946	>0.9	Good fit
2	Relative	Normal fit (NFI)	0.822	>0.9	Acceptable
		Incremental fit (IFI)	1.012	>0.9	Good fit
		Comparative fit (CFI)	1.000	>0.9	Good fit

loading factor of 6.14. But the environment has a negative effect on the labour productivity. Since the p value of the estimates (non-standardized) is all greater than significance level 0.05 for the factor loadings, the estimates do account for large variance and hence cannot be supported. The results of the hypothesis are as follows.

H1: Rejected
H2: Failed to reject
H3: Failed to reject.

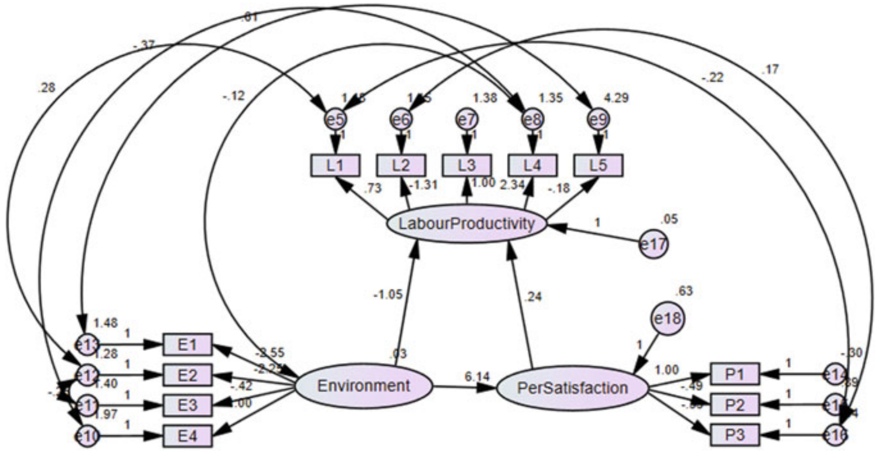


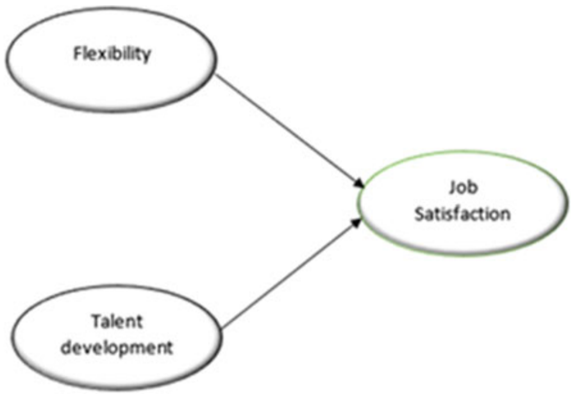
Fig. 4.46 SEM model with estimated parameters for productivity analysis

4.3.4 Application 4: SEM for Job Satisfaction and HR Policies

Problem Description

Many companies have relied on the human resources, and they possess in the past, present and in the expected future as well. As the number of companies and the population increased, a great need for the training of the employees was created and a human resource division was created. Also, every company has been striving to retain the employees, and job satisfaction is one of the key criteria for the employees to retain in an organization or a company. The conceptual model is presented in Fig. 4.47.

Fig. 4.47 Conceptual model for HR policies



So, a model has been created to understand the relationship between the job satisfaction and the complementing human resource policy in a company. The figure shown depicts the representation of the conceptual model and interrelations of the job satisfaction, talent development procedure from the human resource and flexibility in work culture. In this model, talent development and flexibility aspects are directly related to job satisfaction. The problem is to analyse whether this conceptual/theoretical model is accurate or not and to modify the system so that the theoretical model is apt for drawing conclusions. Therefore, two hypotheses have been postulated.

H1: Good talent development policies have a positive impact on job satisfaction.
H2: Flexible work culture has a positive impact on the job satisfaction.

Structural Model

Since the job satisfaction, talent development and flexibility cannot be measured directly, they are labelled as latent variables. Therefore, these variables can be identified or understood using suitable identifiers, also called measured variables. The identifiers or the measured variables for the job satisfaction, talent development and flexibility are listed in Table 4.7. The data for the identifiers are collected using a questionnaire with a rating scale from 1 to 5 where a score of 1 indicates "strongly disagree" and a score of 5 indicates "strongly agree". The data collected consist of 100 values or sample size of 100.

The structural model of this problem is depicted in Fig. 4.48.

According to the procedure of structural model or drawing a path, the latent variables are mentioned in an oval. The identifiers are represented using a rectangle. The error terms are represented using a circle. The single-headed arrow indicates correlation of the causal and the caused variable.

The structural model alone does not give all the info as the measured variables do have a contribution from the errors. Hence, the complete model is shown in Fig. 4.49.

The above model represents the factors responsible for errors as well.

Table 4.7 Latent variables for HR policies

S. No.	Latent variables	Identifiers	
1	Job satisfaction	S1	The productivity of the employee
		S2	The commitment of the employee
		S3	The punctuality of the employee
2	Flexibility	F1	The location of the company
		F2	The working hours
3	Talent development	T1	Training offered
		T2	Promotion in the posts
		T3	Salary

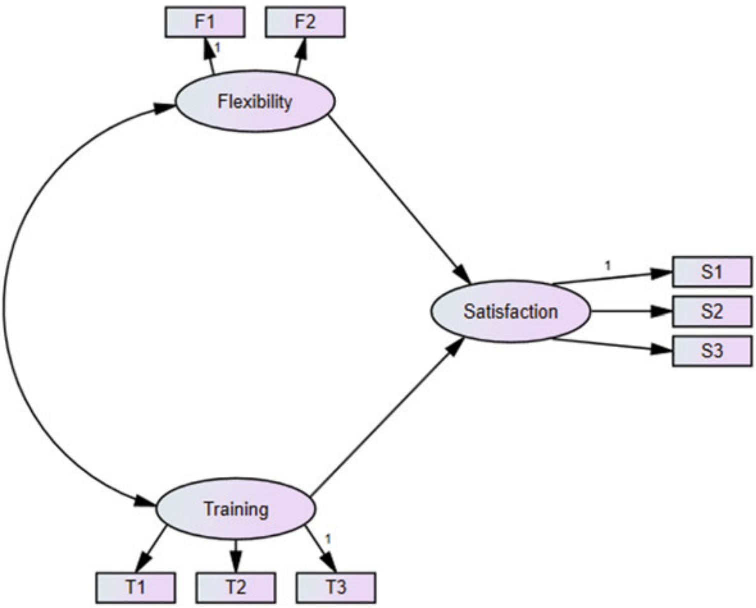


Fig. 4.48 Structural model of HR policies

The assumptions made in this problem are:

9. The data set obtained has a normal distribution.
10. The problem has multivariate normal data.

Analysis

The theoretical model has to be analysed for its appropriateness. There are five steps to analyse the structural equation model developed. They are:

16. Model specification
17. Model identification
18. Parameter estimation
19. Assessment of model fit
20. Modification of the model.

The software used for this analysis is IBM SPSS AMOS v25.

10. ***Model specification***:

The considered model is represented previously. The variables in the model and their types are specified explicitly here. The variables—flexibility and training—are exogenous variables which are independent. The variable—job satisfaction—is endogenous which is dependent on other variables.

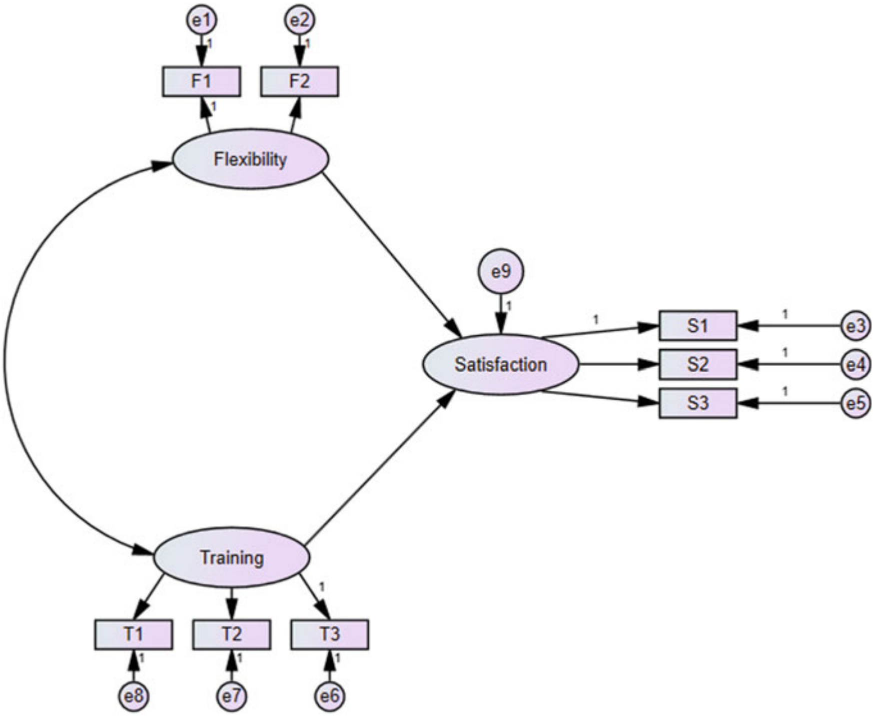


Fig. 4.49 Complete model of HR policies

11. ***Model identification***:

This is done to check whether the unknown parameters can be estimated using the number of knowns or not. If "*s*" is the number of measured variables, the number of known parameters is $\frac{1}{2}(s + 1) * s$. In the model shown above, the number of known parameters is 78. The parameters to be estimated are 3 latent variable variance, 9 error variance and 5 loading factors. This accounts to 17 unknown parameters. The degree of freedom $78 - 17 = 61$.

12. ***Estimation of parameters, model fit and modification***:

The parameters are estimated by using a covariance matrix by the software programme. It is estimated for a fit model only. The initial assessment revealed the following output (Table 4.8).

RMR, GFI, RMSEA
The present model does not have any modification.

The estimates of the unknown parameters:
The regression weights or the load factors are shown in Table 4.9.
Estimates for covariances (Table 4.10).

Table 4.8 Estimated parameters for HR policies

Model	RMR	GFI	AGFI	PGFI
Default model	0.032	0.729	0.426	0.344
Saturated model	0.000	1.000		
Independence model	1.197	0.146	−0.097	0.114

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	0.924	0.874	0.933	0.888	0.932
Saturated model	1.000		1.000		1.000
Independence model	0.000	0.000	0.000	0.000	0.000

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	0.026	0.221	0.304	0.000
Independence model	0.782	0.751	0.814	0.000

Table 4.9 Regression weights for HR policies

			Estimate	S.E.	C.R.	*P*	Label
Satisfaction	←	Training	0.941	0.118	7.953	***	par_6
Satisfaction	←	Flexibility	0.500	0.115	4.337	***	par_7
F1	←	Flexibility	1.000				
F2	←	Flexibility	1.436	0.088	16.393	***	par_1
S1	←	Satisfaction	1.000				
S2	←	Satisfaction	0.957	0.037	26.043	***	par_2
S3	←	Satisfaction	0.927	0.050	18.554	***	par_3
T3	←	Training	1.000				
T2	←	Training	1.501	0.065	22.940	***	par_4
T1	←	Training	1.356	0.059	23.086	***	par_5

Table 4.10 Covariances of HR policies

			Estimate	S.E.	C.R.	*P*	Label
Flexibility	↔	Training	0.788	0.118	6.681	***	par_8

Estimates for the variances (Table 4.11).
Estimates for total effects (Table 4.12).

Assessing the model fitness with fitness indices:
There are two types of indices to check the model fit. They are absolute indices and relative indices. The indices calculated by the software are shown in Table 4.13.

Table 4.11 Estimates of variances for HR policies

	Estimate	S.E.	C.R.	*P*	Label
Flexibility	0.723	0.121	5.997	***	par_9
Training	0.796	0.126	6.315	***	par_10
e9	−0.048	0.009	−5.640	***	par_11
e1	0.133	0.021	6.243	***	par_12
e2	0.304	0.048	6.372	***	par_13
e3	0.120	0.017	6.879	***	par_14
e4	0.104	0.015	6.824	***	par_15
e5	0.293	0.040	7.328	***	par_16
e6	0.097	0.013	7.512	***	par_17
e7	0.125	0.017	7.283	***	par_18
e8	0.099	0.014	7.259	***	par_19

Table 4.12 Estimates of total effects for HR policies

	Training	Flexibility	Satisfaction
Satisfaction	0.668	0.338	0.000
T1	0.968	0.000	0.000
T2	0.967	0.000	0.000
T3	0.944	0.000	0.000
S3	0.606	0.307	0.907
S2	0.645	0.327	0.966
S1	0.644	0.326	0.964
F2	0.000	0.911	0.000
F1	0.000	0.919	0.000

Table 4.13 Fit indices for HR policies

S. No.	Type	Index	Actual value	Preferred value	Conclusion
1	Absolute	Chi-square (χ^2)	0.00	>0.05	Not fit
		Root mean square error of approximation (RMSEA)	0.026	<0.05	Good fit
		Goodness of fit (GFI)	0.729	>0.9	Acceptable
2	Relative	Normal fit (NFI)	0.924	>0.9	Good fit
		Incremental fit (IFI)	0.933	>0.9	Good fit
		Comparative fit (CFI)	0.932	>0.9	Good fit

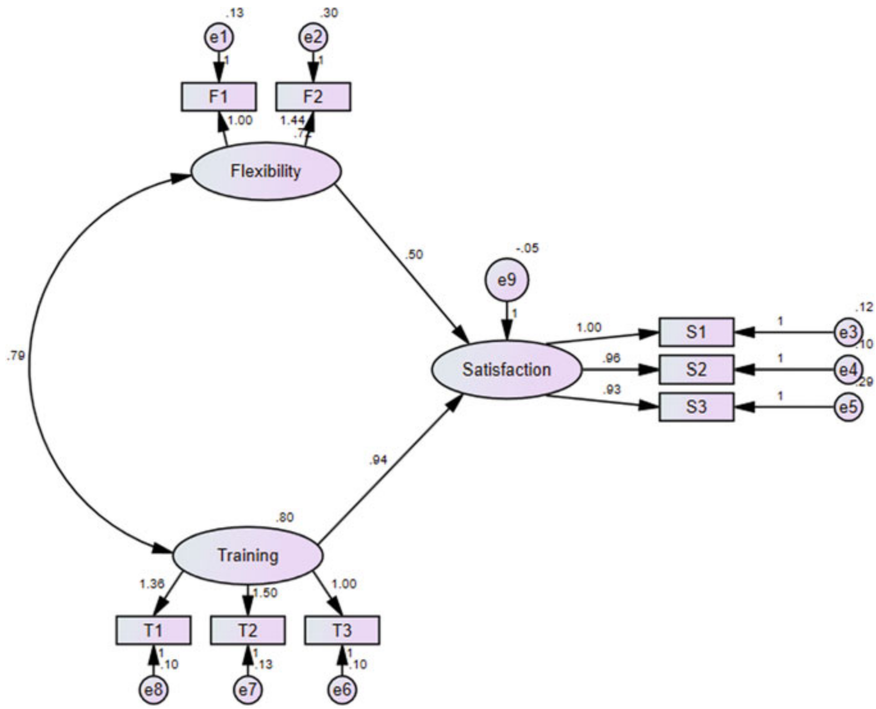


Fig. 4.50 SEM model with estimate parameters

The model with all the estimates will be represented as shown in Fig. 4.50.

Results and Conclusion

As per the initial model, the hypothesis of the modeller was that talent development and flexibility positively influence job satisfaction. After the result, loading factors indicate that the talent development and flexibility positively influence the job satisfaction. Since the p value of the estimates (non-standardized) is all lesser than significance level 0.05 for the factor loadings, the estimates do not account for large variance and hence can be supported. Finally, the hypothesis results are as follows:

H1: Accept
H2: Accept.

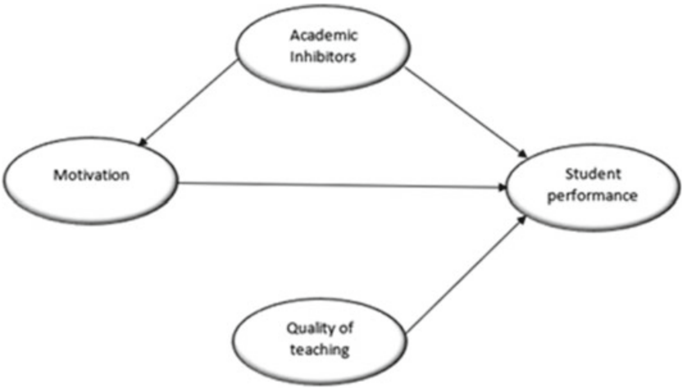


Fig. 4.51 Conceptual model for teaching pedagogy

4.3.5 Application 5: SEM for Student Performance and Teaching Pedagogy Model

Problem Description

A student's academic performance is one of the major success factors for the schools or any educational institution. The performance of the student is dependent on many variables which can be from within the student and the mode of teaching. Figure 4.51 depicts the represent of the conceptual model and interrelations of the student performance, academic inhibitors, motivation and quality of teaching which involves the pedagogy of the teacher.

In this model, quality of teaching, motivation and academic inhibitors are influencing the student performance. The problem is to analyse whether this conceptual/theoretical model is accurate or not and to modify the system so that the theoretical model is apt for drawing conclusions. Therefore, four hypotheses have been postulated.

H1: Academic inhibitors has a negative impact on student performance.
H2: Academic inhibitors has a negative impact on motivation.
H3: Quality of teaching has a positive impact on student performance.
H4: Great motivation has positive impact on the student performance.

Structural Model

The student performance, motivation, academic inhibitors and quality of teaching are labelled as latent variables. Therefore, these variables can be identified or understood using suitable identifiers, also called measured variables. The identifiers or the measured variables for the latent variables are shown in Table 4.14. The data for the identifiers are collected using a questionnaire with a rating scale from 1 to 5 where a score of 1 indicates "strongly disagree" and a score of 5 indicates "strongly agree". The data collected consist of 100 values or sample size of 100.

Table 4.14 Latent variables for teaching pedagogy model

S. No.	Latent variables	Identifiers	
1	Student performance	S1	Marks scored by the student
		S2	Awards received by the student
		S3	Activities involved by the student
2	Academic inhibitors	A1	Non-academic commitments
		A2	Desire to enjoy
3	Quality of teaching	Q1	Inspiration from the teacher
		Q2	Connecting with the students
		Q3	Engagement with the students
4	Motivation	M1	Definite purpose to study
		M2	Competitive environment
		M3	Proper academic planning

The structural model of this problem is depicted in Fig. 4.52.

According to the procedure of structural model or drawing a path, the latent variables are mentioned in an oval. The identifiers are represented using a rectangle. The error terms are represented using a circle. The single-headed arrow indicates correlation of the causal and the caused variable. The structural model alone does not give all the information as the measured variables do have a contribution from the errors. Hence, the complete model is shown in Fig. 4.53.

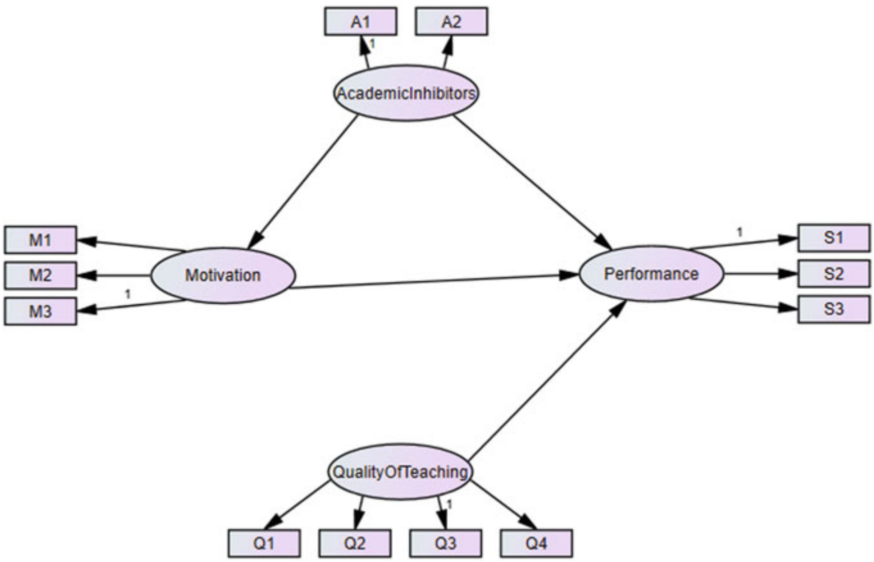


Fig. 4.52 Structural model of teaching pedagogy

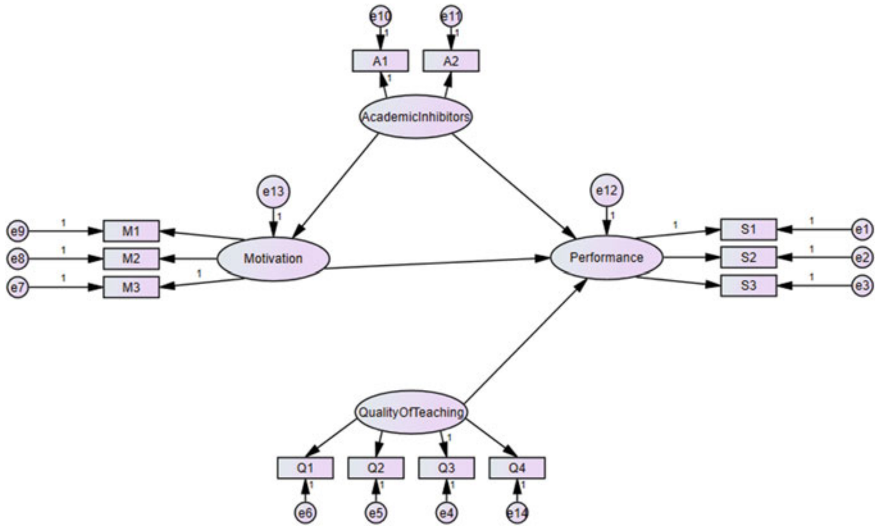


Fig. 4.53 Complete model of teaching pedagogy

The above model represents the factors responsible for errors as well.

The assumptions made in this problem are:

11. The data set obtained has a normal distribution.
12. The problem has multivariate normal data.

Analysis

The theoretical model has to be analysed for its appropriateness. There are five steps to analyse the structural equation model developed. They are:

21. Model specification
22. Model identification
23. Parameter estimation
24. Assessment of model fit
25. Modification of the model.

The software used for this analysis is IBM SPSS AMOS v25.

13. ***Model specification***:

The considered model is represented previously. The variables in the model and their types are specified explicitly here. The variables—inhibitors and quality of teaching—are exogenous variables which are independent. The variables—performance and motivation—are endogenous which is dependent on other variables.

Table 4.15 Estimated parameters for teaching pedagogy

Model	RMR	GFI	AGFI	PGFI
Default model	0.547	0.442	0.111	0.277
Saturated model	0.000	1.000		
Independence model	1.312	0.122	−0.037	0.103

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	0.663	0.546	0.675	0.559	0.673
Saturated model	1.000		1.000		1.000
Independence model	0.000	0.000	0.000	0.000	0.000

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	0.433	0.409	0.457	0.000
Independence model	0.652	0.631	0.672	0.000

14. ***Model identification***:

This is done to check whether the unknown parameters can be estimated using the number of knowns or not. If "*s*" is the number of measured variables, the number of known parameters is ½($s + 1$) * s. In the model shown above, the number of known parameters is 153. The parameters to be estimated are 4 latent variable variance, 14 error variance and 12 loading factors. This accounts to 38 unknown parameters. The degree of freedom $153 - 30 = 123$.

15. ***Estimation of parameters, model fit and modification***:

The parameters are estimated by using a covariance matrix by the software programme. It is estimated for a fit model only. The initial assessment revealed in Table 4.15.

RMR, GFI, RMSEA

The root mean square error of approximation (RMSEA) has a value 0.433 which is more than the prescribed threshold 0.05. The goodness of fit index (GFI) has a value of 0.442 which had to be more than 0.9. So, these values indicate that the model is not fit.

But, the software programme suggests a modification index which can increase the model fitness.

The estimation, assessing model fitness and modification are iterative in AMOS.

After adding a covariance relation between suggested variables, the output of the model fitness analysis after the change is shown in Table 4.16.

After this modification, the model looks as shown in Fig. 4.54.

The estimates of the unknown parameters:
The regression weights or the load factors are calculated in Table 4.17.
Estimates for covariances (Table 4.18).
Estimates for the variances (Table 4.19).
Estimates for total effects (Table 4.20).

Table 4.16 Model fitness analysis for teaching pedagogy model

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	0.177	0.902	0.399	0.369
Saturated model	0.000	1.000		
Independence model	1.312	0.122	−0.037	0.103

Baseline comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	0.932	0.806	0.914	0.825	0.908
Saturated model	1.000		1.000		1.000
Independence model	0.000	0.000	0.000	0.000	0.000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	0.027	0.247	0.299	0.000
Independence model	0.652	0.631	0.672	0.000

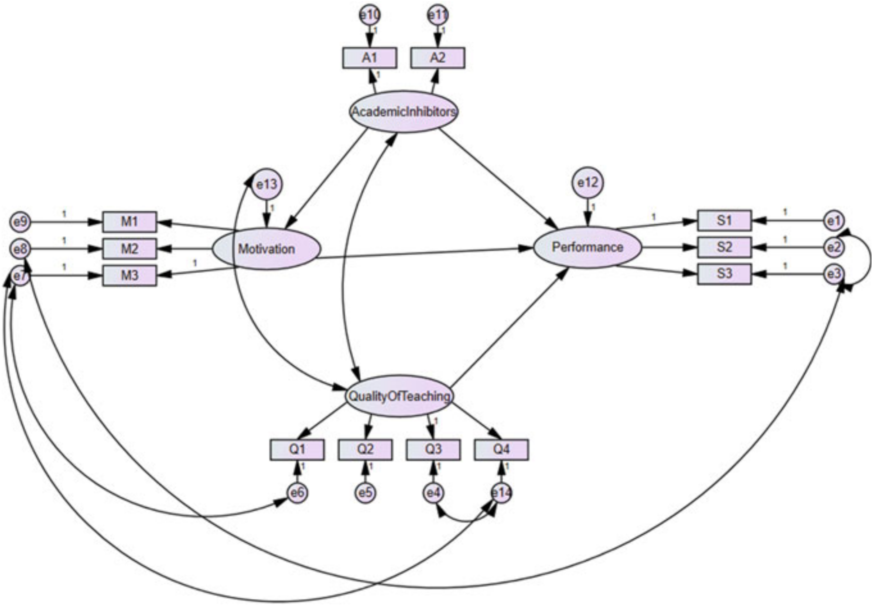


Fig. 4.54 Modified structural model for teaching pedagogy

Table 4.17 Regression weights for teaching pedagogy model

			Estimate	S.E.	C.R.	*P*	Label
Motivation	←	Academic inhibitors	0.388	0.047	8.306	***	par_11
Performance	←	Quality of teaching	−0.264	0.434	−0.608	0.543	par_8
Performance	←	Academic inhibitors	−0.014	0.029	−0.490	0.624	par_9
Performance	←	Motivation	1.834	0.717	2.558	0.011	par_10
S1	←	Performance	1.000				
S2	←	Performance	1.355	0.106	12.761	***	par_1
S3	←	Performance	1.293	0.110	11.712	***	par_2
Q3	←	Quality of teaching	1.000				
Q2	←	Quality of teaching	0.867	0.028	31.467	***	par_3
Q1	←	Quality of teaching	0.855	0.038	22.519	***	par_4
M3	←	Motivation	1.000				
M2	←	Motivation	1.434	0.076	18.830	***	par_5
M1	←	Motivation	1.449	0.067	21.720	***	par_6
A1	←	Academic inhibitors	1.000				
A2	←	Academic inhibitors	0.809	0.046	17.664	***	par_7
Q4	←	Quality of teaching	0.742	0.032	23.078	***	par_13

Table 4.18 Covariances for teaching pedagogy model

			Estimate	S.E.	C.R.	*P*	Label
Quality of teaching	↔	Academic inhibitors	1.277	0.223	5.715	***	par_12
e13	↔	Quality of teaching	0.610	0.091	6.738	***	par_14
e2	↔	e3	0.975	0.142	6.858	***	par_15
e3	↔	e8	0.129	0.023	5.621	***	par_16
e7	↔	e14	0.035	0.011	3.300	***	par_17
e6	↔	e7	0.097	0.019	5.102	***	par_18
e4	↔	e14	−0.044	0.008	−5.235	***	par_19

Table 4.19 Estimates for variances for teaching pedagogy model

	Estimate	S.E.	C.R.	*P*	Label
Quality of teaching	1.748	0.252	6.934	***	par_20
Academic inhibitors	2.195	0.280	7.848	***	par_21
e13	0.377	0.063	5.962	***	par_22
e12	−0.180	0.033	−5.498	***	par_23
e1	0.310	0.048	6.407	***	par_24
e2	0.974	0.143	6.821	***	par_25
e3	1.201	0.151	7.942	***	par_26
e4	0.025	0.008	3.029	0.002	par_27
e5	0.111	0.016	6.764	***	par_28
e6	0.230	0.033	6.941	***	par_29
e7	0.129	0.018	7.145	***	par_30
e8	0.157	0.023	6.720	***	par_31
e9	0.041	0.008	5.164	***	par_32
e10	−0.284	0.088	−3.236	0.001	par_33
e11	0.409	0.077	5.294	***	par_34
e14	0.097	0.016	6.180	***	par_35

Table 4.20 Estimates of total effects for teaching pedagogy model

	Academic inhibitors	Quality of teaching	Motivation	Performance
Motivation	0.388	0.000	0.000	0.000
Performance	0.698	−0.264	1.834	0.000
Q4	0.000	0.742	0.000	0.000
A2	0.809	0.000	0.000	0.000
A1	1.000	0.000	0.000	0.000
M1	0.563	0.000	1.449	0.000
M2	0.557	0.000	1.434	0.000
M3	0.388	0.000	1.000	0.000
Q1	0.000	0.855	0.000	0.000
Q2	0.000	0.867	0.000	0.000
Q3	0.000	1.000	0.000	0.000
S3	0.903	−0.341	2.371	1.293
S2	0.946	−0.358	2.485	1.355
S1	0.698	−0.264	1.834	1.000

Assessing the model fitness with fitness indices:
There are two types of indices to check the model fit. They are absolute indices and relative indices. The indices calculated by the software are shown in Table 4.21.

The model with all the estimates will be represented as shown in Fig. 4.55.

Results and Conclusion
As per the initial model, academic inhibitors have a negative impact on performance, academic inhibitors have a negative impact on motivation, quality of teaching has a positive impact on performance, and great motivation has positive impact on the

Table 4.21 Fit indices for teaching pedagogy

S. No.	Type	Index	Actual value	Preferred value	Conclusion
1	Absolute	Chi-square (χ^2)	0	>0.05	Not fit
		Root mean square error of approximation (RMSEA)	0.027	<0.05	Good fit
		Goodness of fit (GFI)	0.902	>0.9	Good fit
2	Relative	Normal fit (NFI)	0.932	>0.9	Good fit
		Incremental fit (IFI)	0.914	>0.9	Good fit
		Comparative fit (CFI)	0.908	>0.9	Good fit

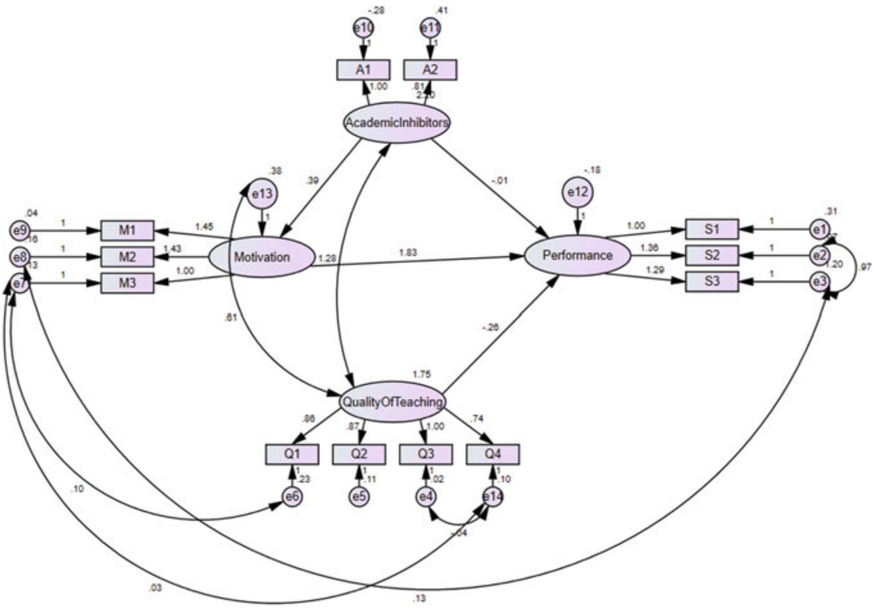


Fig. 4.55 SEM model with estimated parameters for teaching pedagogy

performance. After the result, loading factors indicate that the motivation positively influences the performance, quality of teaching, and inhibitors negatively influence the performance. Since the p value of the estimates (non-standardized) is all lesser than significance level 0.05 for the factor loadings, the estimates do not account for large variance and hence can be supported. Finally, the hypothesis results are as follows:

H1: Accept
H2: Reject
H3: Reject
H4: Accept.

Chapter 5
Applications of Structural Equation Modelling with R

5.1 History of R Software

R software typically is a computer language and run-time environment developed by Ross Ihaka and Robert Gentleman based on older programming language S. The first version of R was released in 1990s and several people have joined this development project for its improvement. In 1995, R is launched as "open-sourced" software and anyone can use and modify it for their problem environment. In the 2000, the first version of R (1.0) was released for the public. R includes a wide range of standard statistical and graphical analyses tools and also provides a large number of user-developed extension packages to extend greater set of capabilities to the users. This software can be obtained without any licensing fees and it can also be distributed to others. The other softwares like SPSS AMOS and SAS also provides very good user-friendliness for conducting SEM analysis but the high-licensing cost makes their acceptability for an individual researcher less attractive.

R is a user-friendly computer package which provides an environment for statistical computing and graphics. It is a GNU project (Free software package provided by MIT) which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S. There are select key differences between S and R software.

R is equipped with many statistical techniques such as linear and nonlinear modelling, classical statistical tests, time-series analysis, classification and clustering and graphical techniques. Typically, S language is considered as the vehicle of choice for research in statistical methodology, and R intends to provide an open source route to participation in that activity.

R compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and MacOS. This extends an outstanding capability in producing well-designed publication-quality plots.

J. J. Thakkar, *Structural Equation Modelling*, Studies in Systems, Decision and Control 285, https://doi.org/10.1007/978-981-15-3793-6_5

R is a software facility for data manipulation, calculation and graphical display. The key features of R include:

- an effective data handling and storage facility
- a suite of operators for calculations on arrays, in particular matrices
- a large, coherent and integrated collection of intermediate tools for data analysis
- graphical facilities for data analysis and display either on-screen or on hardcopy and
- a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

R, like S, is developed around a true computer language, which permits users to add additional functionality by defining new functions. Much of the system is itself written in the R dialect of S, which makes it easy for users to follow the algorithmic choices made. For computationally intensive tasks, C, C++ and Fortran code can be linked and called at run time. Advanced users can write C code to manipulate R objects directly.

It is necessary for an analyst to understand that R is not just a statistics system but it is an environment within which statistical techniques are implemented. R can be extended (easily) via *packages*. There are about eight packages supplied with the R distribution and many more are available through the CRAN family of Internet sites covering a very wide range of modern statistics.

Lavaan Package for SEM in R:

Structural equation modelling (SEM) is a growing field and extensively used by many the researchers in various applications of engineering, medical science, social sciences, humanities and law. The last three decades have observed the development of many software packages for solving structural equation modelling which are primarily closedsource and/or commercial. The **R** package **lavaan** has been developed to provide applied researchers, teachers and statisticians, a free, fully open source, but commercial quality package for latent variable modelling.

Structural equation models are solved using many softwares like LISREL, Amos, R, etc. Here, three applications of SEM have been solved using R with the interface of RStudio. Few R packages are needed as prerequisites which include "lavaan" and "semPlot". A brief explanation of the problem and the syntax for solving the problem in R is explained. The two problems are illustrated here with the application of R software. The problems are:

1. Student performance and teaching pedagogy model
2. Productivity model.

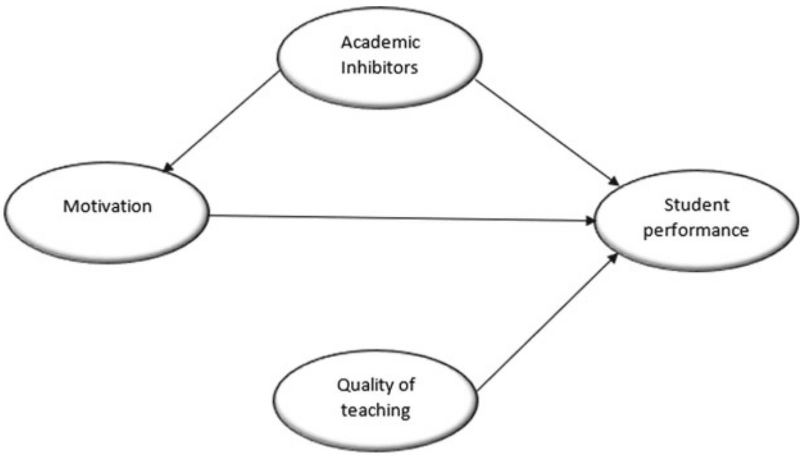
5.2 Step-by-Step Procedure for Conducting SEM in R

Step 1: Open RStudio.
Step 2: Import and load “lavaan” library.
Step 3: Import and load “semPlot” package.
Step 4: Define and build the model with all the constructs and latent variables.
Step 6: Feed the dataset and solve the model by using “sem” function.
Step 7: Display the (graphical) model.
Step 8: Display the estimates and indices.
Step 9: Interpret the results.

5.3 Illustrative Applications of SEM in R

5.3.1 Application 1: Student Performance and Teaching Pedagogy Model

A student’s academic performance is one of the major success factors for the schools or any educational institution. The performance of the student is dependent on many variables which can be from within the student and the mode of teaching. The figure shown depicts the representation of the conceptual model and interrelations of the student performance, academic inhibitors, motivation and quality of teaching which involves the pedagogy of the teacher.



In this model, quality of teaching, motivation and academic inhibitors have been considered as the key constructs for evaluating student performance. The problem is to analyse whether this conceptual/theoretical model is accurate or not and to modify

Table 5.1 Identifiers for latent variables

S. No.	Latent variables	Identifiers	
1	Student performance (P)	S1	Marks scored by the student
		S2	Awards received by the student
		S3	Activities involved by the student
2	Academic inhibitors (A)	A1	Non-academic commitments
		A2	Desire to enjoy
3	Quality of teaching (Q)	Q1	Inspiration from the teacher
		Q2	Connecting with the students
		Q3	Engagement with the students
4	Motivation (M)	M1	Definite purpose to study
		M2	Competitive environment
		M3	Proper academic planning

the system so that the theoretical model is apt for drawing conclusions. Therefore, four hypotheses have been postulated.

H1: Academic inhibitors have a negative impact on student performance.
H2: Academic inhibitors have a negative impact on motivation.
H3: Quality of teaching has a positive impact on student performance.
H4: Motivation has a positive impact on the student performance.

Table 5.1 denotes the identifiers for the latent variables.

The syntax for solving this problem using R is explained as follows:

```
library(lavaan)                #loading the library lavaan
library(semPlot)               #Loading the package semPlot
stdt='                         #Defining the model
A=~A1+A2
P=~S1+S2+S3
Q=~Q1+Q2+Q3+Q4
M=~M1+M2+M3
P~A+M+Q
M~A
'
stdt                           #Recalling the model
std=sem(stdt,data=Stdnt_data)  #Solving the model using "sem"
                               function(Stdnt_data is the
                               database filename which was
                               imported)
semPaths(std)                  #Displaying the model in R
summary(std)                   #Displaying the solved estimates
fitmeasures(std)               #Displaying the fit indices
modindices(std)                #Displaying the modification
indices
```

Table 5.2 Fit indices

S. No.	Type	Index	Actual value	Preferred value	Conclusion
1	Absolute	Chi-square (χ^2)	0	>0.05	Not fit
		Root mean square error of approximation (RMSEA)	0.412	<0.05	Not fit
		Goodness of fit (GFI)	0.404	>0.9	Not fit
2	Relative	Normal fit (NFI)	0.693	>0.9	Not fit
		Incremental fit (IFI)	0.705	>0.9	Not fit
		Comparative fit (CFI)	0.703	>0.9	Not fit

After executing the above syntax, the fit indices shown by the results are as in Table 5.2.

After viewing the above process, modify the model as suggested by the modification indices.

After the modification, the model is modified as follows:

```
stdt='
A=~A1+A2
P=~S1+S2+S3
Q=~Q1+Q2+Q3+Q4
M=~M1+M2+M3
P~A+M+Q
M~A
Q4 ~~  M3
Q3 ~~  Q4
Q1 ~~  M3
S3 ~~  Q1
S2 ~~  S3
A2 ~~  M2
A1 ~~  M2
A1 ~~  S3
A1 ~~  A2
A1 ~~  S2
A2 ~~  S3
S1 ~~  S2
S1 ~~  S3
'
```

After the modification, the indices are displayed in Table 5.3.

Since most of the indices are acceptable, the model can be accepted.

Table 5.3 Fit indices after modification

S. No.	Type	Index	Actual value	Preferred value	Conclusion
1	Absolute	Chi-square (χ^2)	0	>0.05	Not fit
		Root mean square error of approximation (RMSEA)	0.276	<0.05	Not fit
		Goodness of fit (GFI)	0.680	>0.9	Not fit
2	Relative	Normal fit (NFI)	0.899	>0.9	Fit
		Incremental fit (IFI)	0.903	>0.9	Good fit
		Comparative fit (CFI)	0.902	>0.9	Good fit

The final loading factor matrix is displayed as follows:

```
Regressions:
                                    Estimate
 P ~
   A                                  0.965
   M                                 -0.935
   Q                                  1.224
 M ~
   A                                  1.258
```

From the above regressions:
The hypothesis has the following results:

H1: Reject
H2: Reject
H3: Accept
H4: Reject

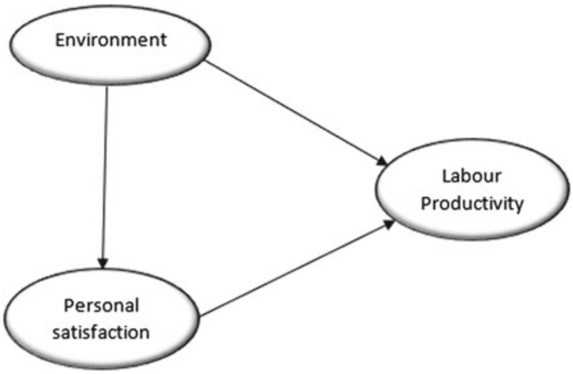
5.3.2 *Application 2: Productivity Model*

It has become a usual practice in almost all the sectors of profession to improve the productivity. In this line, productivity can be extended to the entities like labour, machine, etc. Productivity can be defined by the ratio of output to the input and multiplied by 100 whenever the expectation is in terms of percentage.

The current problem deals with a model involving labour productivity. It can be observed that labour productivity depends on the environment of working and personal satisfaction of the labourers. Also, environment might influence the personal satisfaction to some extent.

Table 5.4 Identifiers for latent variables for Productivity model

S. No.	Latent variables	Identifiers	
1	Labour productivity (L)	L1	Quantity of input consumption
		L2	Sales generated
		L3	Discipline among the workers
		L4	Number of tasks done
		L5	Interest developed in the work
2	Work environment (W)	E1	Lighting conditions in the workplace
		E2	Noise
		E3	Safety measures taken
		E4	Co-workers relationship
3	Personal satisfaction (P)	P1	Commuting time spent
		P2	Health condition
		P3	Family commitments



The figure shown depicts the representation of the conceptual model and interrelations of the labour productivity along with its contributing factors viz. environment in which the labourer works and personal satisfaction of the labourer.

In this model, environment has a direct correlation with the labour productivity as well as indirect correlation through personal satisfaction. The personal satisfaction has a direct effect on the labour productivity. The problem is to analyse whether this conceptual/theoretical model is accurate or not and to modify the system so that the theoretical model is apt for drawing conclusions.

Therefore, the following hypotheses are postulated.

H1: Good environment has a positive influence on the labour productivity.
H2: Good environment has a positive influence on the personal satisfaction.
H3: Personal satisfaction has a positive influence on the labour productivity (Table 5.4).

The syntax for solving this problem using R is explained as follows:

```
library(lavaan)                   #loading the library lavaan
library(semPlot)                  #Loading the package semPlot
prod='                            #Defining the model
L=~L1+L2+L3+L4
W=~E1+E2+E3+E4
P=~P1+P2+P3
L~W+P
P~W
'
prod                              #Recalling the model
prods=sem(prod,data=Prod_data)    #Solving the model using "sem"
                                  function(Prod_data is the database
                                  filename which was imported)
semPaths(prods)                   #Displaying the model in R
summary(prods)                    #Displaying the solved estimates
fitmeasures(prods)                #Displaying the fit indices
modindices(prods)                 #Displaying the modification
indices
```

After executing the above syntax, the fit indices shown by the results are as follows in Table 5.5.

After viewing the above process, modify the model as suggested by the modification indices.

Table 5.5 Fit indices for productivity model

S. No.	Type	Index	Actual value	Preferred value	Conclusion
1	Absolute	Chi-square (χ^2)	0.00	>0.05	Not fit
		Root mean square error of approximation (RMSEA)	0.088	<0.05	Not fit
		Goodness of fit (GFI)	0.897	>0.9	Acceptable
2	Relative	Normal fit (NFI)	0.641	>0.9	Acceptable
		Incremental fit (IFI)	0.791	>0.9	Acceptable
		Comparative fit (CFI)	0.774	>0.9	Acceptable

Table 5.6 Fit indices after modification for productivity model

S. No.	Type	Index	Actual value	Preferred value	Conclusion
1	Absolute	Chi-square (χ^2)	0.360	>0.05	Good fit
		Root mean square error of approximation (RMSEA)	0.025	<0.05	Good fit
		Goodness of fit (GFI)	0.953	>0.9	Good fit
2	Relative	Normal fit (NFI)	0.846	>0.9	Acceptable
		Incremental fit (IFI)	0.988	>0.9	Good fit
		Comparative fit (CFI)	0.987	>0.9	Good fit

After the modification, the model is modified as follows:

```
prod='
L=~L1+L2+L3+L4
W=~E1+E2+E3+E4
P=~P1+P2+P3
L~W+P
P~1*W
E2~~P1
E1~~E2
L2~~P3
L1~~E2
P1~~P2
E2~~P2
L1~~P1
L1~~P3
L3~~E1
L3~~P2
'
```

After the modification, the indices are displayed in Table 5.6.
Since, most of the indices are acceptable, the model can be accepted.
The final loading factor matrix is displayed as follows:

```
Regressions:
                                  Estimate
  L ~
    W                               0.101
    P                               0.131
  P ~
    W                               1.000
```

From the above regressions:
The hypothesis has the following results:

H1: Accept
H2: Accept
H3: Accept

Chapter 6
Applications of SEM and FAQs

6.1 Applications of SEM

SEM has widespread of applications in business, psychology and industry. Traditionally, the cause and effect relationship is traditionally solved using some regression technique when the independent and dependent variables are continuous or measurable. However, the variables or constructs are influenced by several other variables. For example, employee motivation is influenced by HR policy, working environment, salary structure and many other factors which involve cause and effect relationships. SEM is a powerful technique to solve such problems. The summary of area-specific application of SEM is given in Table 6.1.

6.2 Frequently Asked Questions (FAQs) on Structural Equation Modelling (SEM)

Question 1:
I am a first-year MBA student. I need to know about capabilities of SEM analysis and the software available for conducting such analysis.
Answer 1:
SEM is a multivariate analysis technique which combines the use of regression analysis and factor analysis for testing the hypotheses and confirming the relationships among observed and latent variables. The widely used software for SEM are IBM SPSS AMOS, R and SAS.

SEM is considered as an advanced multivariate analysis technique which offers three unique advantages:

(i) explicit assessment of measurement error
(ii) estimation of latent (unobserved) variables via observed variables; and
(iii) model testing in which a structure can be imposed and assessed as to fit of the data.

J. J. Thakkar, *Structural Equation Modelling*, Studies in Systems, Decision and Control 285, https://doi.org/10.1007/978-981-15-3793-6_6

Table 6.1 Area-specific application of SEM

S. No.	References	What is being investigated?	What are the variables?
Operations Management			
1	Vázquez-Bustelo et al. (2007)	Analyses agile manufacturing in Spain and study whether it is a critical factor for success in different industries	• Manufacturing strength (I/P: cost, quality flexibility, delivery, service, environment; O/P: labour productivity, customer loyalty, new product development success, sales volume, return on assets, responsiveness to changes in competitive conditions) • Agile manufacturing (agile human resources, agile technologies, value chain integration, concurrent engineering, knowledge management) • Turbulent environment (dynamism, hostility)
2	Sin et al. (2015)	Structural equation modelling on knowledge creation in Six Sigma DMAIC project and its impact on organizational performance	• Socialization • Externalization • Combination • Internalization • Knowledge • Six Sigma project success • Organizational performance
3	Zanjirchi et al. (2017)	Risk-agility interactive model: a new look at agility drivers	• **Organizational agility** (Flexibility, accountability, change culture, speed, Integration, quality, competency, human resources) • **Supply chain risk** Product risk (quality, cost, process) Environmental risk (human resources disorder, transportation risk, sovereign risk, the risk of natural disasters) Knowledge risk (control, recognition) ICT risk (database risk, software risk) Customer risk (demand risk, prediction risk, delay risk) The risk of customer satisfaction (risk of after-sales services, risk of customer relationship management) Supplier risk (purchase risk, order fulfilment risk) Distributor risk (availability risk, risk of new distribution channels)

(continued)

Table 6.1 (continued)

S. No.	References	What is being investigated?	What are the variables?
4	Van Poucke et al. (2016)	Enhancing cost savings through early involvement of purchasing professionals in sourcing projects: Bayesian estimation of a structural equation model	• Internal customer satisfaction • Early purchasing involvement • Cost savings • Strategic impact
5	Lu et al. (2007)	Application of structural equation modelling to evaluate the intention of shippers to use Internet services in liner shipping	• Security protection • Perceived usefulness • Perceived ease of use • Use intention
Marketing Management			
1	Kharouf et al. (2014)	Building trust by signalling trustworthiness in service retail	• Trustworthiness (integrity, benevolence, competence, value alignment, consistency, communication, behavioural loyalty, attitudinal loyalty) • Trustworthiness • Trust
2	Mishra et al. (2018)	Technology readiness of teenagers: a consumer socialization perspective	• **Motivators and inhibitors as separate dimensions** (Age, concept-oriented communication, informative media, informative peer, normative peer, socio-oriented communication, motivators, inhibitors) • **Overall technology readiness** (Age, concept-oriented communication, informative media, informative peer, normative peer, socio-oriented communication, technology readiness)
3	Singh (2015)	Modelling passengers' future behavioural intentions in airline industry using SEM	• Customer services with empathy • In-flight services and facilities • Convenience and promptness with reliability with reliability • Airline perceived image (IMG) • Perceived value (VAL) • Passenger satisfaction (SAT) brand image

(continued)

Table 6.1 (continued)

S. No.	References	What is being investigated?	What are the variables?
4	Jin et al. (2012)	How do individual personality traits (D) influence perceived satisfaction with service for college students (C) in a casual restaurant setting (I)?: the CID framework	• Consumer relationship proneness • Need for variety • Confidence benefit • Social benefit • Special treatment • Satisfaction
5	Jain et al. (2018)	Examining consumer-brand relationships on social media platforms	• Involvement • Satisfaction • Commitment • Brand trust • Brand loyalty • WOM
Organizational Behaviour			
1	Tan et al. (2015)	Linkage between knowledge management and manufacturing performance: a structural equation model ng approach	• Knowledge resources (human capital, knowledge and information capital, Intellectual property) • KM processes (knowledge acquisition, knowledge creation and generation, knowledge utilization and application, knowledge storing and updating, knowledge sharing and transferring, knowledge protection); KM factors (culture, management leadership and support, organizational infrastructure and technology, strategy)
2	Azam et al. (2013)	Structural equation modelling (SEM)-based trust analysis of Muslim consumers in the collective religion affiliation model in e-commerce	• Initial trust on Website • Consumer religion-centrism • Religious commitment • Intention to purchase from same religious Websites
3	Prasertratana et al. (2014)	Distributed leadership: structural equation model	• Collaboration (shared resources, shared decision making, mutuality) • Creativity (originality, fluency, flexibility, elaboration) • Trust factors (competence, honesty, benevolence, reliability, openness)
4	Pokharel and Choi (2015)	Exploring the relationships between the learning organization and organizational performance	• Continuous learning • Inquiry and dialogue • Collaboration and team learning • System to capture learning • Empower employee • Connect the organization • Strategic leadership

(continued)

Table 6.1 (continued)

S. No.	References	What is being investigated?	What are the variables?
Human Resource Development			
1	de Beer et al. (2016)	Job insecurity, career opportunities, discrimination and turnover intention in post-apartheid South Africa: examples of informative hypothesis testing	• Job insecurity • Career paths • Discrimination
2	Froese and Xiao (2012)	Work values, job satisfaction and organizational commitment in China	• Individualism • Job autonomy • Willingness to take risks • Appraisal satisfaction • Pay satisfaction • Money orientation • Organizational commitment
3	Panda and Rath (2017)	The effect of human IT capability on organizational agility: an empirical analysis	• Business functional capability • Interpersonal management capability • Technology management capability • Organizational sensing agility • Organizational responding agility • IT spending
4	Tummers et al. (2015)	Connecting HRM and change management: the importance of proactivity and vitality	• Training and development • Feedback • Job autonomy • Participation • Teamwork • Proactivity • Vitality
5	Shih and Tsai (2016)	The effects of knowledge management capabilities on perceived school effectiveness in career and technical education	• School effectiveness (administrative efficiency, teaching effectiveness, research outcome) • KM process capabilities (acquisition, storage, sharing, application) • KM enabler capabilities (structure, culture, IT support)
Psychology			
1	Rezaei and Ghazanfari (2016)	The role of childhood trauma, early maladaptive schemas, emotional schemas and experimental avoidance on depression: a structural equation modelling	• Childhood trauma (emotional abuse, physical abuse, emotional neglect, physical neglect) • Disconnection (defectiveness/shame, abandonment, emotional deprivation, isolation, mistrust) • Negative emotional schemas (rigid emotional schemas, negative beliefs about emotion) • Experimental avoidance • Depression

(continued)

Table 6.1 (continued)

S. No.	References	What is being investigated?	What are the variables?
2	Li and Jiang (2018)	Social exclusion, sense of school belonging and mental health of migrant children in China: a structural equation modelling analysis	• Social exclusion (exclusion in leisure activity, exclusion in study, exclusion in friendship) • Sense of school belonging (class atmosphere, school participation, sense of closeness)
3	Lee and Kim (2018)	A structural equation model on Korean Adolescents' excessive use of smartphones	• Impulsiveness (family function, peer relationship), self-esteem (family function, peer relationship) • Friends' support (family function, peer relationship) • Excessive use of smartphones (family function, peer relationship, impulsiveness, self-esteem, friends support)
4	Khoreva and Tenhiälä (2016)	Gender differences in reactions to injustice	• Procedural justice • Pay inequity • Internal comparison • External comparison • Knowledge of pay • Organizational commitment
5	Wang et al. (2018)	Psychological pathway to suicidal ideation among people living with HIV/AIDS in China: a structural equation model	• Perceived stigma (anticipated stigma, negative self-image, public attitude, disclosure concern) • Depression • Social support • Self-esteem

Most multivariate techniques do not accommodate an error component explicitly in the model, whereas SEM models estimate these error variance parameters for both independent and dependent variables. In addition, SEM helps to estimate the latent variables from observed variables. Finally, a researcher can test a fully developed model against the data using SEM as a conceptual or theoretical structure or model and can be evaluated for fit of the sample data.

Question 2:

I am conducting a market research and interested in evaluating the impact of various marketing strategies on sales. Can I conduct such analysis using SEM? Do I have to purchase software for this as it is a self-motivated research?

Answer 2:

SEM is extensively used for market research and certainly you can analyse the impact of various marketing strategies such as price discount, advertisement, buy-back options on customer behaviour and sales for the various potential markets. There is no need to purchase a software for conducting SEM analysis. You can register online for using a free trial version of IBM SPSS software or you can use R software which is an open-source free software.

Question 3:

I am acting as General Manager (HR) in a multinational Company. How SEM can help me to reveal greater insights on HR policies?

Answer 3:

SEM offers a huge potential to investigate into the measured observable and latent constructs. The outcomes of HR policies are usually evaluated in terms of employee satisfaction, employee turnover (attrition rate), improvement in human productivity, employee absenteeism level, etc. These constructs can be defined by the number of indicators. The constructs have set of linear relationships with HR policies and may have causal relationships among them. This makes it a fit case for conducting SEM analysis. I would advise you to select the key constructs and related indicators influenced by HR policies for the context of your company and conduct the SEM analysis as per the step-by-step procedure described in this book.

Question 4:

I am just a beginner in the domain of multivariate statistical analysis. Can you explain in the simplest term, how can I appreciate the use of SEM?

Answer 4:

It is very simple. Just think about a situation, if I run for one hour at the speed of 7 km/h, I may burn 500 kcal. This can be represented as run–burn calorie. The relationship can easily be represented in the form of a path diagram.

Now just think about a case in which select variables are not directly measurable (typically called as latent variables). For example, "Erratic behaviour of the student" is a latent construct which cannot be directly measured. It might be occurring because of the various symptoms such as (i) student might have some problem in home and hence he has an erratic behaviour; (ii) student might be sick which results in erratic behaviour; (iii) student is not comfortable with teaching pedagogy and hence feels irritable, etc. In this case, our latent variable is "Erratic behaviour of the student" which is the cause of all these symptoms. We can hypothesize the relationships between symptoms and cause in terms of path diagram. The necessary data may be collected which may represent a particular relationship. We then try to investigate that "Is it possible that this pattern of relationships I have proposed in my model can explain the pattern of relationships I've found in my data". If we get a positive answer then our proposed model (hypothesized relationships) are right, otherwise we need to once again verify our assumptions about the proposed relationships.

Question 5:

I am confused about the goodness of regression and SEM analysis. Which one is better?

Answer 5:

Regression intends to define the relationship between multiple correlated observed independent (predictor) variables and one dependent variable (called observed variable). In comparison to this, SEM has much wider capability to analyse complex theories. Typical advantages of SEM include:

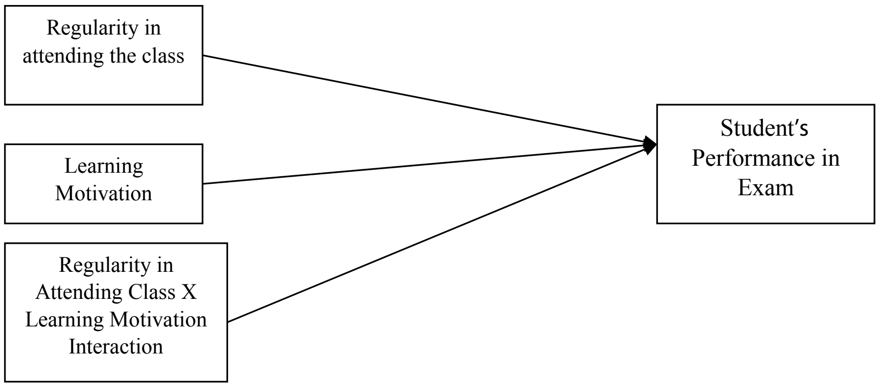
1. It accommodates multiple observed variables to measure the latent variable and builds a measurement model for each of your predictor- and dependent variables. It enables an analyst to control measurement error and get more precise estimates for the regression coefficients between the latent variables.
2. It permits the development of a model for testing complex theories which has dependent variables acts as predictor variables for other dependent variables. This allows you to examine indirect effects and mediation structures.
3. It is primarily used as confirmatory factor analysis and helps to translate a good theoretical model into a statistical model. If the pattern prevailing in the data set cannot explain the hypothesized relationships then the estimated parameters can be interpreted and SEM will not represent the structure/pattern prevailing in the data.

Question 6:
SEM analysis involves different kinds of variables. Can you provide a simple and easy to understand explanation of these variables?
Answer 6:
The variables involved in SEM analysis can be explained as follows:

- **Exogenous variables**: They are not influenced by other variables in a model. For example, we consider two factors that can impact the health of a person such as "food" and "exercise". We can see that as such there is no relationship between these two variables and hence are called exogenous variables in the model.
- **Endogenous variables**: They are influenced by other variables in a model. We understand that "food" and "exercise" can impact the person's health, and hence, "Persons Health" is considered as a endogenous variable in the model.
- **Manifest variable**: It is a directly observed and measured variable, also know an indicator variable. For example, "Food" and "Exercise" both can be measured in terms of appropriate units and hence are called manifest variable. In SEM model, manifest variables are examined through *path analysis*.
- **Latent variable**: It cannot be directly measured. The "factors" in a factor analysis are latent variables. For example, "awareness on health" is also to be investigated in terms of its impact on "Health of a person", and then this can be indirectly assessed through a "self-discipline" adopted by a person in regularity of "food" and "exercise". It is an indirect measurement, and hence, "self-discipline" is a latent variable. Latent variables increase the complexity of a structural equation model because one needs to take into account all of the questionnaire items and measured responses that are used to quantify the "factor" or latent variable.

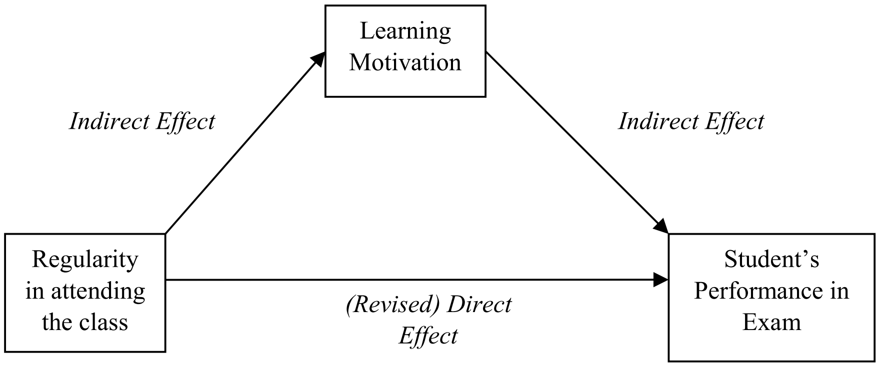
Question 7:
What are moderation and mediation effects in SEM?
Answer 7:
Moderation refers to a situation that includes three or more variables to examine the association between two variables under the influence of third variable (set at different levels). This is similar to examining an interaction effect in ANOVA. For example, regularity in attending class and student performance in exam may differ

at different levels of learning motivation (i.e., this is the definition of an interaction). This means it is very much possible that regularity in attending class may have different degrees of influence on student's performance when it is assessed at different levels of learning motivation. This indicates a two-way interaction between regularity in attending classes and student's performance in exam. Figure shows a conceptual diagram of moderation. This diagram shows that there are three direct effects that are hypothesized to cause changes in student performance in exam—a main effect of regularity in attending the class, a main effect of learning motivation and an interaction effect of regularity in attending class and learning motivation.



Example of Moderation

Mediation refers to a situation that includes three or more variables with causal relationships between all three variables. It is necessary to understand that this is different from moderation. In case of moderation, we have considered three separate things in the model that may cause a change in student's performance in exam: regularity in attending the class, learning motivation and an interaction effect of regularity in attending class and learning motivation. Mediation describes a much different relationship that is generally more complex. A mediation relationship considers a direct effect between an independent variable and a dependent variable. There are also indirect effects between an independent variable and a mediator variable and between a mediator variable and a dependent variable. The main difference in mediation compared to moderation is that it allows for causal relationships between regularity in attending the classes and learning motivation and student's performance in exam to be expressed. Suppose, if we assume that learning motivation is not included in the model and we are just interested to examine the direct effect of regularity in attending the class and student's performance in exam, this will give us a measure of the direct effect by using regression or ANOVA. When a model includes learning motivation as a mediator, that direct effect will change as a result of decomposing the causal process into indirect effects of regularity in attending the class on learning motivation and effect of learning motivation on student's performance in exam. The degree to which the direct effect changes as a result of including the mediating variable of learning motivation is referred to as the mediational effect. Testing for mediation involves running a series of regression analysis for all of the causal pathways and some method of estimating a change in direct effect.



Example of Mediation

The moderation and mediational models are considered as the foundation of SEM analysis and can be considered as simple structural equation models themselves. Therefore, it is very important to understand how to analyse such models to understand more complex structural equation models that include latent variables.

Question 8:

I am a final year mechanical engineering student. Can I use SEM for doing my project in the domain of production engineering?

Yes. The capabilities of SEM analysis make it a widely popular technique for analysing various issues in different domains. For example, an organizational productivity is governed by many measured and latent variables such as government policies, quality of raw material, technology, motivation of workers, standardization of the processes. The variables may have linear and causal relationships between them. This can be considered as a candidate for SEM analysis. If we think about an example typically in the domain of manufacturing processes, then the impact of many factors on quality of manufacturing process can be considered as the problem for investigation through SEM analysis.

Question 9:

Can missing data in SEM analysis like other statistical analysis create a problem? How do we handle missing data situation in SEM analysis?

Yes. Like other statistical methods, in SEM also missing data can reduce the statistical power of the analysis and results in poor investigation. The most widely used technique for missing data estimation is maximum likelihood methods. Majority of the software packages are equipped with this method.

Question 10:
I am investigating a problem in social science. What should be considered as an adequate sample size for SEM analysis?
Answer 10:
As a rule of thumb, SEM examines the complex structures explaining a particular theory and requires sample sizes of at least 200 to examine basic models. However, if you are dealing with a higher model complexity including more number of observed and latent variables then you need to consider more than 250 samples for achieving a better statistical power. Two popular assumptions are that you need more than 200 observations, or at least 50 more than 8 times the number of variables in the model. A larger sample size is always desired for SEM.
Question 11:
What are the key precautions in the use of SEM technique?
Answer 11:
There are six precautions in the use of SEM technique.

1. SEM is used as a confirmatory technique and a researcher must specify a full model a priori and test that model based on the sample and variables included in your measurements. This demands knowing the number of parameters one need to estimate including the associated covariances, path coefficients, and variances. It is also important to deduce relationships from prior theory or experience between the variables in SEM model.
2. SEM is used to model complex relationships between multivariate data, and hence, sample size is an important issue. It is recommended that a researcher should consider 200 observations or at least 50 more than 8 times the number of variables in the model. A larger sample size is always desired for SEM.
3. Like other multivariate statistical methodologies, most of the estimation techniques used in SEM require multivariate normality. Your data need to be examined for univariate and multivariate outliers. Transformations on the variables can be made. However, there are some estimation methods that do not require normality.
4. SEM techniques only look at first-order (linear) relationships between variables. Linear relationships can be explored by creating bivariate scatter plots for all of your variables. If a relationship between two variables seems to be quadratic, then power transformation can be applied.
5. There is a key concern for multicollinearity among the independent variables for manifest variables. Most programmes will inspect the determinant of a section of your covariance matrix or the whole covariance matrix. A very small determinant may be indicative of extreme multicollinearity.
6. It is expected that in SEM analysis, the residuals of the covariances need to be small and centred about zero. Some goodness of fit tests (like the Lagrange multiplier test) remain robust against highly deviated residuals or non-normal residuals.

Question 11:
Why SEM has received significant admiration in research?
Answer 11:
The use of SEM is admired for theory testing for its unique capabilities. This includes:

- SEM examines whether the proposed model produces a population covariance matrix that is consistent with the sample covariance matrix. SEM can help to answer many questions when we are trying to test a priori model with set of linear and causal relationships.
- SEM draws the attention of a research on whether the proposed theoretical model is adequate or not using goodness of fit indices. The method can also be used to compare multiple theoretical models derived for explaining a particular phenomenon.
- SEM enables a researcher to understand the amount of variance in the dependent variables (DVs)—both manifest and latent DVs—are accounted for by the IVs.
- SEM can also be used to investigate the group differences. A researcher can derive a separate SEM model for different groups and compare the results.

Summary of Key Points

- SEM is not just a "statistical technique" but it integrates a number of different multivariate techniques into one model fitting framework. SEM integrates the various techniques such as measurement theory, factor (latent variable) analysis, path analysis, regression and simultaneous equations.
- SEM confirms the causality among constructs and variables based on the evidence of systematic covariation and helps the researcher to justify that relationships are not false.
- SEM is a typical form of graphical modelling and therefore a system in which relationship can be represented in either graphical or mathematical form.
- SEM is also known as covariance structure analysis, Analysis of moment structures, analysis of linear structural relationships (LISREL), causal modelling.
- SEM helps to understand to what extent we are able to explain the structure of relationships among the various exogenous and endogenous variables leading to the final effect under investigation. Usually, we deal with linear structure but lots of advancements are there to accommodate nonlinear structure also.
- SEM is considered as path analysis using latent variables.
- SEM is a framework for building and evaluating multivariate hypotheses about multiple processes. It is not dependent on a particular estimation method.
- A model for SEM analysis must be derived from a sound underlying theory, and this should include the development of measurement model specification and structural model specification.
- SEM seeks to progress knowledge through cumulative learning. It advances the knowledge from exploratory to confirmatory/hypothesis testing.
- SEM is considered to be a decision process which involves six steps (Hair et al., 2012):
  - ***Step 1***: Definition of individual constructs (what items are to be used as measured variables?).
  - ***Step 2***: Development of measurement model and path diagram.

J. J. Thakkar, *Structural Equation Modelling*, Studies in Systems, Decision and Control 285, https://doi.org/10.1007/978-981-15-3793-6

- ***Step 3***: Design of a study to produce empirical results (Check the adequacy of sample size and select an appropriate estimation method and missing data approach).
- ***Step 4***: Assessment of the measurement model validity (Assess GOF and construct validity of measurement model). If measurement model is found valid, then proceed to stages 5 and 6 otherwise refine measures and design a new study.
- ***Step 5***: Convert measurement model into a structural model.
- ***Step 6***: Assessment of the structural model validity (Assess GOF and significance, direction and size of structural parameter estimates). If the structural model is valid, then draw the detailed conclusions and recommendations otherwise refine model and test with new data.

- Underlying parameters in SEM
  - Regression coefficients
  - Variances and covariances.
- A hypothesized model is used to estimate these parameters for the population (assuming the model is true)
  - The parameter estimates are used to create a hypothesized variance/covariance matrix
  - The sample VC matrix is compared to the estimated VC.
- Combination of factor analysis and regression
  - Continuous and discrete predictors and outcomes
  - Relationships among measured or latent variables.
- Direct link between *Path Diagrams* and equations and fit statistics.
- SEM models include both measurement and path models.
- Principal component analysis versus factor analysis:
  - In principal components analysis, the goal is to explain as much of the total variance in the variables as possible. The goal in factor analysis is to explain the covariances or correlations between the variables.
  - Use principal components analysis to reduce the data into a smaller number of components. Use factor analysis to understand what constructs underlie the data.
  - In principal components analysis, the components are calculated as linear combinations of the original variables. In factor analysis, the original variables are defined as linear combinations of the factors.
- Factor loadings represent how much a factor explains a variable in factor analysis.
- A communality is the extent to which an item correlates with **all** other items. Higher communalities are better. If communalities for a particular variable are low (between 0.0 and 0.4), then that variable may struggle to load significantly on any factor.

- Relationship between factor model and degree of freedom (DOF): If DOF = 0 (unique solution); DOF > 0 (multiple solutions); DOF < 0 (factor model is indetermined).
- In SEM analysis, first develop measurement model (which is basically confirmatory factor analysis) and then based on that develop structural model. If structural model is valid, then only draw detailed conclusions.
- Multivariate regression versus path model: There is no interrelationship among independent variables in multivariate regression model and also among dependent variables. There is only a relationship between dependent and independent variables. In path model, there is interrelationship among (within) dependent variables and among (within) the independent variables.
- **Endogenous (dependent latent) versus exogenous (independent latent) variable**: Exogenous variables are the one which are independent latent variables and not affected by any other variable in the model (usually presented on the left-hand side of the model) whereas others are endogenous variables which are either affected by exogenous variables or endogenous variables.
- **Goodness of fit indices**: Measurement model and structural model both should fit to the data—must be checked using goodness of fit indices like—chi-square, root mean square residual, goodness of fit Index (GFI), normed fit index (NFI), comparative fit index (CFI), incremental fit index (IFI).
- **Absolute fit versus relative fit indices**: Relative fit indices explain the goodness of fit of the present model with respect to many other models possible whereas absolute fit indices explain the goodness of fit for the model itself (under consideration).
- Three most important issues for measurement model and structural model are: (i) model identification; (ii) parameter estimation; (iii) model adequacy test.
- For the identification of measurement model, two conditions should be satisfied: (i) order condition and (ii) rank condition. Order condition is bare minimum to be satisfied and it satisfies the necessity and rank condition satisfies the sufficiency.
- Order condition: Number of parameters to be identified must be less than number of non-redundant elements in the covariance matrix. Preferably model should be over identified. Once model identification is done, you are ready for model parameter estimation.
- Important assumption in measurement model or confirmatory factor analysis is that the covariance between manifest variable or latent variable and error term is zero.
- The output of measurement model is relationship between exogenous and endogenous variables. This is utilized as an input to structural model.
- Structural model is basically path analysis. We conduct path analysis between exogenous and endogenous variables whose relationships are determined based on measurement model. We are not focusing on manifest variables.
- Every endogenous variable (right side of the structural model) may have direct effect or indirect effect. We have to identify the total effect (direct and indirect) on

the issue/problem/phenomena under investigation. This helps to rank or prioritize the variables.

- The widely used software for SEM are: LISREL, MPlus, EQS, Amos, Calis, Mx, SEPATH, Tetrad, R, Stata.

YouTube Links

S. No.	Topic	YouTube link	Time (in min)
1	Structural equation modelling: what is it and what can we use it for?	https://www.youtube.com/watch?v=eKkESdyMG9w by Professor Patric Sturgis, Department of Social Statistics and Demography at the University of Southampton, published by National Centre for Research Methods (NCRM)	25.32
2	Introduction to structural equation modelling	https://www.youtube.com/watch?v=Hukuz1uNdjs by Professor J. Maiti, IIT Kharagpur published by National Programme for Technology Enhanced Learning (NPTEL)	55.45
3	Key ideas, terms and concepts in SEM	https://www.youtube.com/watch?v=NOWdrfQVWAI by Professor Patric Sturgis, Department of Social Statistics and Demography at the University of Southampton, published by National Centre for Research Methods (NCRM)	41.20
4	Why use a structural equation model?	https://www.youtube.com/watch?v= -m4ag3WQcCw by Mr. Dan, Curran-Bauer Analytics	11.28
5	Structural equation modelling using AMOS	https://www.youtube.com/watch?v=fwrp0D7T1UI by G N Satish Kumar, My easy statistics	12.47

J. J. Thakkar, *Structural Equation Modelling*, Studies in Systems, Decision and Control 285, https://doi.org/10.1007/978-981-15-3793-6

Glossary of Key SEM Terminologies

Absolute fit indices It intends to examine overall goodness of fit for both the structural and measurement models jointly. It does not compare the assumed model with a null model (incremental fit measure) or attempts to adjust the number of parameters in the estimated model (parsimonious fit measure). It is a direct measure of how well the model specified by the researcher produces the observed data. The SEM software packages compute multiple absolute fit indices such as: chi-square statistic, goodness of fit (GFI), root means square residual (RMSR), standardized root mean residual (SRMR), root mean square error of approximation (RMSEA).

Coefficient of determination It is the fraction of variation explained by an equation of a model.

Composite reliability The proportion of the variance of a variable which is not due to measurement error.

Confirmatory factor analysis It is a factor analysis which finds the loading factors to describe the correlation between the variables which are defined a priori. The factors restricted.

Construct It is a broad concept which is defined conceptually with a theoretical meaning. Constructs may be abstract and need not to be directly observable. Examples: Intelligence or customer satisfaction.

Constructs are considered as unobservable which are represented by set of variables. A typical mathematical relationship among variables represents a construct.

Exogenous constructs are latent and equivalent to independent variables. These are constructs that are determined outside of the model.

Exogenous constructs are latent equivalent to dependent variables.

Direct effect The effect of a variable on another due to the direct relationship between them in the SEM.

Endogenous variable A variable is called endogenous if all the arrows point into it and also called dependent variable. It is represented by a variate of dependent variables.

Error Variance leftover after prediction of a measured variable.

J. J. Thakkar, *Structural Equation Modelling*, Studies in Systems, Decision and Control 285, https://doi.org/10.1007/978-981-15-3793-6

Exogenous variable A variable is called exogenous if the paths originate from it and also called independent variable. These are the constructs which are determined by factors outside of the model.

Exploratory factor analysis It is a factor analysis which finds the loading factors to describe the correlation between the variables. In this type, the factors are not restricted and all are connected to each other.

Goodness of fit index (GFI) It measures how well a specified model reproduces the covariance matrix among the observed variables. This means to what extent the similarity exists between the observed and estimated covariance matrices.

Imputation It tries to estimate the values of missing data based on the valid values of the other variables. This is accomplished by identifying relationships with the valid sample and using this identified relationship for estimating the value of a missing observation.

Incremental fit indices These are goodness of fit indices used to examine how well a specified model fits relative to an another baseline model. The baseline model is considered as the null model which assumes that all the measured variables are unrelated to each other. It supports the other goodness of fit indices such as absolute fit and parsimonious fit. The SEM programmes compute multiple incremental fit indices such as normed fit index (NFI), comparative fit index(CFI), Tucker Lewis index(TLI), relative noncentrality index (RNI).

Indirect effect The effect of a variable on another due to an intermediate variable between them in the SEM.

Instrumental variable A variable which introduces exogenous variability into an endogenous variability.

Latent variable A variable in the model which cannot be measured. It can also be called factor or unmeasured variable.

Loading factors It is a factor which indicates the effect of one variable on another.

Measurement model It indicates how measurement variables together represent constructs.

Measurement model The part of the model that relates indicators to latent factors. It is the factor analytic part of SEM.

Model fit The ability of an overidentified model to reproduce the correlation or covariance matrix of the variables.

Model identification Identification of the model is to know whether the solution exists for a model or not. A model is said to be identified if there exists a unique solution for all parameters.

Modification index It is a score test for adding paths where none occur.

Multicollinearity It explains to what extent a particular construct can be explained by the other constructs in the analysis. An increase in multicollinearity raises the complexity in interpretation as a particular construct has many interrelationships with other variables and it is difficult to determine the effect of this construct in the analysis.

Observed variable A variable in the model which is present in the data set and measured. It is also called as an indicating variable for a latent variable.

Over fit model It is a model which is very much exactly related to a particular set of data and fails to fit additional data of the system. It has a limitation in predicting the future reliably. It includes more parameters (typically known as noise factors or residual variations) than the parameters necessary to fit the data and represent an underlying model structure.

Path diagram A graphical representation of the structural equation model using ellipses, circles, rectangles single-headed and double-headed arrows.

Parsimonious fit indices It indicates the overall goodness of fit by establishing relationship between degree of model fit and estimated coefficient. It enables a researcher to correct the model for overfitting by evaluating the parsimony ratio of the model. It complements the other two goodness of fit indices such as absolute fit and incremental fit.

Parsimonious ratio It compares the degree of freedom of specified model and the total number of degrees of freedom available. It indicates to what extent the model utilizes the total number of degrees of freedom available and helps in the examination of overfitting the model with additional relationships that ensure only marginal gains in model fit.

Path analysis It determines the strength of the paths indicated in path diagram using simple bivariate correlations and estimates the relationships in SEM model.

Path diagram It is a graphical and visual representation of all the relationships among the model's constructs. A straight arrow is used to indicate the dependence relationships with an arrow emanating from the predictor variable and the arrow head pointing to the dependent construct or variable. Usually, curved arrows are used to represent correlations between constructs or indicators without any causation.

Path model This is the part of the model that relates variables or factors to one another (prediction). If no factors are in the model then only path model exists between indicators.

Recursivity It is a state where all causal is unidirectional and disturbances are uncorrelated.

Reliability It is used to indicate the internal consistency of a set of indicators of a latent construct on the basis that how highly indicators are interrelated. It indicates an ability of the indicators to measure the same thing (latent construct). The deviation in measurement indicates the degree of error.

Residual It represents the difference between the actual and estimated value for any relationship. Typically, in SEM analysis, it is the difference between the observed and estimated fitted covariance matrices.

Root means square residual (RMSR) It is the error in the prediction of each covariance term which indicates the residual created. RMSR is the square root of the mean of these squared residuals which represents an average of the residuals between individual observed and estimated covariance and variance terms. A standardized value of RMSR is known as standardized root mean residual (SRMR). Lower values of SRMR and RMSR indicate the better fit and higher value represents the poor fit.

Structural model The model in which the parameters are a description of the scenario and also representing causal effects between the variables. It indicates how the constructs are associated with each other. It is also known as theoretical model or causal model.

Structural equation modelling (SEM) It is a multivariate analysis technique which integrates two methodologies such as factor analysis and multiple regression analysis to simultaneously examine series of interrelated dependence relationships among the measured variables and latent constructs (variables) as well as between several latent constructs.

Under fit model It fails to adequately capture the underlying structure of the data which has some parameters missing that are necessary to ensure the correctness of the model. Typically, this occurs while fitting a linear model to nonlinear data.

Variable Actual items that are measured using a survey, observation or some other measurement device. It is created by developing constructs into measurable form. It represents a characteristic which varies and has at least two possible values. Variables are considered as observable items in the sense that we can get a direct measurement of them.

Example: Productivity, age of the students, weight, etc.

Variance and covariance matrix The matric which represents the variance and covariance between two variables. The diagonal elements represent the variance and the other elements represent the covariance between the variables in respective row and column of that element.

Bibliography and Further Reading

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the third Berkeley symposium on mathematical statistics and probability* (Vol. V, pp. 111–150). Berkeley: University of California Press.

Arhonditsis, G. B., Paerl, H. W., Valdes-Weaver, L. M., Stow, C. A., Steinberg, L. J., & Reckhow, K. H. (2007). Application of Bayesian structural equation modeling for examining phytoplankton dynamics in the Neuse River Estuary (North Carolina, USA). *Estuarine, Coastal and Shelf Science, 72*(1–2), 63–80.

Azam, A., Qiang, F., Abbas, S. A., & Ibrahim Abdullah, M. (2013). Structural equation modeling (SEM) based trust analysis of Muslim consumers in the collective religion affiliation model in e-commerce. *Journal of Islamic Marketing, 4*(2), 134–149.

Cudeck, R., Du Toit, S., & Sörbom, D. (Eds.). (2001). *Structural equation modeling: Present and future. A Festschrift in honor of Karl Jöreskog*. Lincolnwood, IL: Scientific Software International.

de Beer, L. T., Rothmann Jr., S., & Pienaar, J. (2016). Job insecurity, career opportunities, discrimination and turnover intention in post-apartheid South Africa: Examples of informative hypothesis testing. *The International Journal of Human Resource Management, 27*(4), 427–439.

Froese, F. J., & Xiao, S. (2012). Work values, job satisfaction and organizational commitment in China. *The International Journal of Human Resource Management, 23*(10), 2144–2162.

Gopal, P. R. C., & Thakkar, J. J. (2016). Sustainable supply chain practices: An empirical investigation on Indian automobile industry. *Production Planning and Control, 27*(1), 49–64.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2012). *Multivariate data analysis* (6th ed.). London: Pearson.

Hair, J. F, Jr., Sarstedt, M., Hopkins, L., & Kuppelwieser, V. G. (2014). Partial least squares structural equation modeling (PLS-SEM). *European Business Review, 26*(2), 106–121.

Hershberger, S. L. (2003). The growth of structural equation modeling: 1994–2001. *Structural Equation Modeling, 10*(1), 35–46.

Jain, N. K., Kamboj, S., Kumar, V., & Rahman, Z. (2018). Examining consumer-brand relationships on social media platforms. *Marketing Intelligence & Planning, 36*(1), 63–78.

Jin, N., Lee, S., & Gopalan, R. (2012). How do individual personality traits (D) influence perceived satisfaction with service for college students (C) in a casual restaurant setting (I)?: The CID framework. *Journal of Hospitality Marketing & Management, 21*(6), 591–616.

Kharouf, H., Lund, D. J., & Sekhon, H. (2014). Building trust by signaling trustworthiness in service retail. *Journal of Services Marketing, 28*(5), 361–373.

Khoreva, V., & Tenhiälä, A. (2016). Gender differences in reactions to injustice. *Journal of Managerial Psychology, 31*(4), 790–804.

Lee, H., & Kim, J. (2018). A structural equation model on Korean Adolescents' excessive use of smartphones. *Asian Nursing Research*.

J. J. Thakkar, *Structural Equation Modelling*, Studies in Systems, Decision and Control 285, https://doi.org/10.1007/978-981-15-3793-6

Li, C., & Jiang, S. (2018). Social exclusion, sense of school belonging and mental health of migrant children in China: A structural equation modeling analysis. *Children and Youth Services Review, 89*, 6–12.

Lu, C. S., Lai, K. H., & Cheng, T. E. (2007). Application of structural equation modeling to evaluate the intention of shippers to use Internet services in liner shipping. *European Journal of Operational Research, 180*(2), 845–867.

Mishra, A., Maheswarappa, S. S., & Colby, C. L. (2018). Technology readiness of teenagers: A consumer socialization perspective. *Journal of Services Marketing*.

Panda, S., & Rath, S. K. (2017). The effect of human IT capability on organizational agility: An empirical analysis. *Management Research Review, 40*(7), 800–820.

Pokharel, M. P., & Ok Choi, S. (2015). Exploring the relationships between the learning organization and organizational performance. *Management Research Review, 38*(2), 126–148.

Prasertratana, S., Sanratana, W., & Somprach, K. O. (2014). Distributed leadership: Structural equation model. *Procedia-Social and Behavioral Sciences, 116*, 1610–1614.

Rezaei, M., & Ghazanfari, F. (2016). The role of childhood trauma, early maladaptive schemas, emotional schemas and experimental avoidance on depression: A structural equation modeling. *Psychiatry Research, 246*, 407–414.

Shih, W. L., & Tsai, C. Y. (2016). The effects of knowledge management capabilities on perceived school effectiveness in career and technical education. *Journal of Knowledge Management, 20*(6), 1373–1392.

Sin, A. B., Zailani, S., Iranmanesh, M., & Ramayah, T. (2015). Structural equation modelling on knowledge creation in Six Sigma DMAIC project and its impact on organizational performance. *International Journal of Production Economics, 168,* 105–117.

Singh, A. K. (2015). Modeling passengers' future behavioral intentions in airline industry using SEM. *Journal of Advances in Management Research, 12*(2), 107–127.

Tan, L. P., & Wong, K. Y. (2015). Linkage between knowledge management and manufacturing performance: A structural equation modeling approach. *Journal of Knowledge Management, 19*(4), 814–835.

Tummers, L., Kruyen, P. M., Vijverberg, D. M., & Voesenek, T. J. (2015). Connecting HRM and change management: The importance of proactivity and vitality. *Journal of Organizational Change Management, 28*(4), 627–640.

Van Poucke, E., Matthyssens, P., & Weeren, A. (2016). Enhancing cost savings through early involvement of purchasing professionals in sourcing projects: Bayesian estimation of a structural equation model. *Journal of Purchasing and Supply Management, 22*(4), 299–310.

Vázquez-Bustelo, D., Avella, L., Fernández, E. (2007). Agility drivers, enablers and outcomes: Empirical test of an integrated agile manufacturing model. *International Journal of Operations & Production Management, 27*(12), 1303–1332.

Wang, W., Wang, Y., Xiao, C., Yao, X., Yang, Y., Yan, H., & Li, S. (2018). Psychological pathway to suicidal ideation among people living with HIV/AIDS in China: A structural equation model. *Psychiatry Research, 260*, 255–261.