Methods of Structuring Business Intelligence

Lung-Hsing Kuo, Hsieh-Hua Yang, Hsueh-Chih Lin, Miao-Kuei Ho, Hung-Jen Yang

***Abstract*—**Business Intelligence (BI) represents the ability to look into the core of a business, in order to fundament the most effective and profitable decisions. An operational BI system sustains daily activities through the following functionalities: real-time informing, secured access to information and easy to use analysis. This change is a natural response of a passing to a new organizational culture of management based on measurable objectives. An operational BI system assumes tracking down trends, problems and other factors as soon as they act, allowing employers to solve them in real time. There is a need to organize business intelligence. A structural business intelligence should be verified based upon theories gained by experiences. Structural equation modeling (SEM) is a versatile statistical modeling tool. Its estimation techniques, modeling capacities, and breadth of applications are expanding rapidly. This study identified some common terminologies. General steps of SEM for verifying BI were discussed along with important considerations in each step. Simple examples are provided to illustrate some of the ideas for structural BI. The intent of this study was to focus on foundational issues to inform readers of the potentials as well as the limitations of verifying BI with SEM.

***Keywords*—**Structure Equation Model, Business Intelligence

I. INTRODUCTION

In this Information Age, corporations have at their disposal enormous amounts of data collected in transactional systems. These systems are designed for the well-organized selection, storage, and retrieval of data, and are vital for businesses to keep track of their relationships.[1-7]

Having data is not the same as having information. The challenge is in deriving answers to business questions from the available data. This wealth of data can yield critical information about a business, so that decision makers at all levels can respond quickly to changes in the business climate.

The information called business intelligence, BI, should be presented in a certain structure and verified, so it can be trusted and operated[8].

Manuscript received February 1, 2013: Revised version received March 25, 2013. This work was supported in part by the Taiwan National Science Council under High Scope Project Grant 102-3113-S-017-003-GJ.

L.H. Kuo is with the National Kaohsiung Normal University, 80201 Taiwan, R.O.C. (e-mail:admi@nknucc.nknu.edu.tw)

H.H.Yang is with the Oriental Institute of Technology, 220 Taiwan, R.O.C. (e-mail:yansnow@gmail.com)

H.C. Lin is with the Taipei College of Maritime Technology, 25172, R.O.C. (e-mail: linhsiaochih@yahoo.com.tw)

M.K. Ho is with the National Kaohsiung Normal University, 80201 Taiwan, R.O.C. (e-mail: hmq0402@gmail.com, phone: 886-7-7172930 ext. 7603,; fax: 886-7-6051206).

H.J. Yang is with the National Kaohsiung Normal University, 80201 Taiwan, R.O.C. (e-mail: hjyang@nknucc.nknu.edu.tw, phone: 886-7-7172930 ext. 7603,; fax: 886-7-6051206).

II. BI FORMULATION

Aggregating data into levels at which patterns can come into view, ordering levels into hierarchies to support drilling down and up through the levels, and using investigative functions such as lag, moving total, and year-to-date are among the techniques used to transform data into information. This information can provide a major boundary in a competitive marketplace.[9]

Business intelligence provides answers to basic questions such as:

- "What are call-center top five questions?"
- "How do a call-center works this year compare to works last year?"
- "What is the 3-month moving average of call-center works?"

Business intelligence can also answer more probing analytical questions such as:

- Why are services down in this region?
- What can we predict for call-center's working load next quarter?
- What factors can we alter to improve the services forecast?
- How will our margins improve if we run this promotion?

Answering these questions requires an analysis of past performance, so that key decision makers can set a course for their businesses that will improve future performance, provide a more competitive edge, and thus enhance profitability. There is a certain way to formulate BI.

Formulating a business intelligence requires careful planning to assure that it meets expectations. These are the basic steps:

- Identify Rationale
- Identify the Data Sources
- Design the Hypothesis Data Model
- Create the Data Store
- Verifying the Model
- Generate the Summary Data
- Prepare the Data for Client Access
- Grant Access Rights
- Distribute the Client Software and Documentation
- Create and Distribute Reports

A. Identify Rationale

It is important to anticipate how end users will analyze the data. By interviewing key users, you can identify the questions that the business intelligence system needs to answer.

One can ask questions such as:

- What information do you have now?
- What additional information do you need?
- How do you want the information presented?

Table 1 Sample requirements might be addressed

Department	Requirements
Board of Directors	Competitive analysis
	Key indicator tracking
	Trend analysis
	Exception reporting
Administrative Analysis and Planning	Investment and acquisitions assessment
	Reorganization analysis
	Long-range planning
	Resource allocation
	Capacity planning
	Human resource planning
Finance Department	Budgeting
	Consolidation
	Variance analysis
	Financial modeling
	Cash management
	Asset liability modeling
	Activity-based management
Sales and Marketing Department	Product profitability
	Customer profiling
	Distribution analysis
	Sales performance and effectiveness

Business requirements can be generated at all levels of an organization. In Table 1., items listed are examples of the requirements one might need to address.

One can find out about the reports and data sources currently available, and what users like and dislike about their current information system. One may also discover what expectations they have about run-time performance.

B. Identify the Data Sources

From the types of questions that end users want answered, one can identify the sources of the data that can provide the answers. The data can be distributed among numerous locations, such as transactional databases and flat files. If the data is not available within your company, then you should discuss whether it is possible to acquire the data or whether end users must modify their expectations.

C. Design the Hypothesized Data Model

The logical data model must support the needs and expectations of your end users. This layer of metadata presents the data in business terms so that users can quickly identify the data they need to use.

For OLAP tools, you define dimensions, measures, and so forth. Then you can map the metadata objects to the physical data sources. For relational tools, you define items, calculations, joins, and so forth using any existing relational data source

There different ways for define a data model, such as dimensions, levels, attributes, hierarchies, cubes, or measures.

D. Create the Data Store

One must deploy the data model as physical objects in the database and load the data from its sources.

For OLAP tools, the data store is an analytic workspace. For relational tools, the data store may be the current OLTP system or a star schema in a data warehouse.

E. Verifying the Model

According the existed data, a statistical test should be conducted for verifying the proposed model. Whether the model is robust ? Whether the model is fitting into the data stored? These questions should be answered before further applying the proposed model.

F. Generate the Summary Data

Business intelligence data is essentially hierarchical, so that data can be summarized at various levels. For performance, some of this data (ideally the data most frequently queried) is summarized and stored as a data maintenance procedure.

In analytic workspaces, summary data is stored in the same analytic workspace objects as the base-level data. In relational schemas, summary data is stored in materialized views.

G. Prepare the Data for Client Access

The client tools query the metadata to find out what data is available, where to get it, and how to present it.

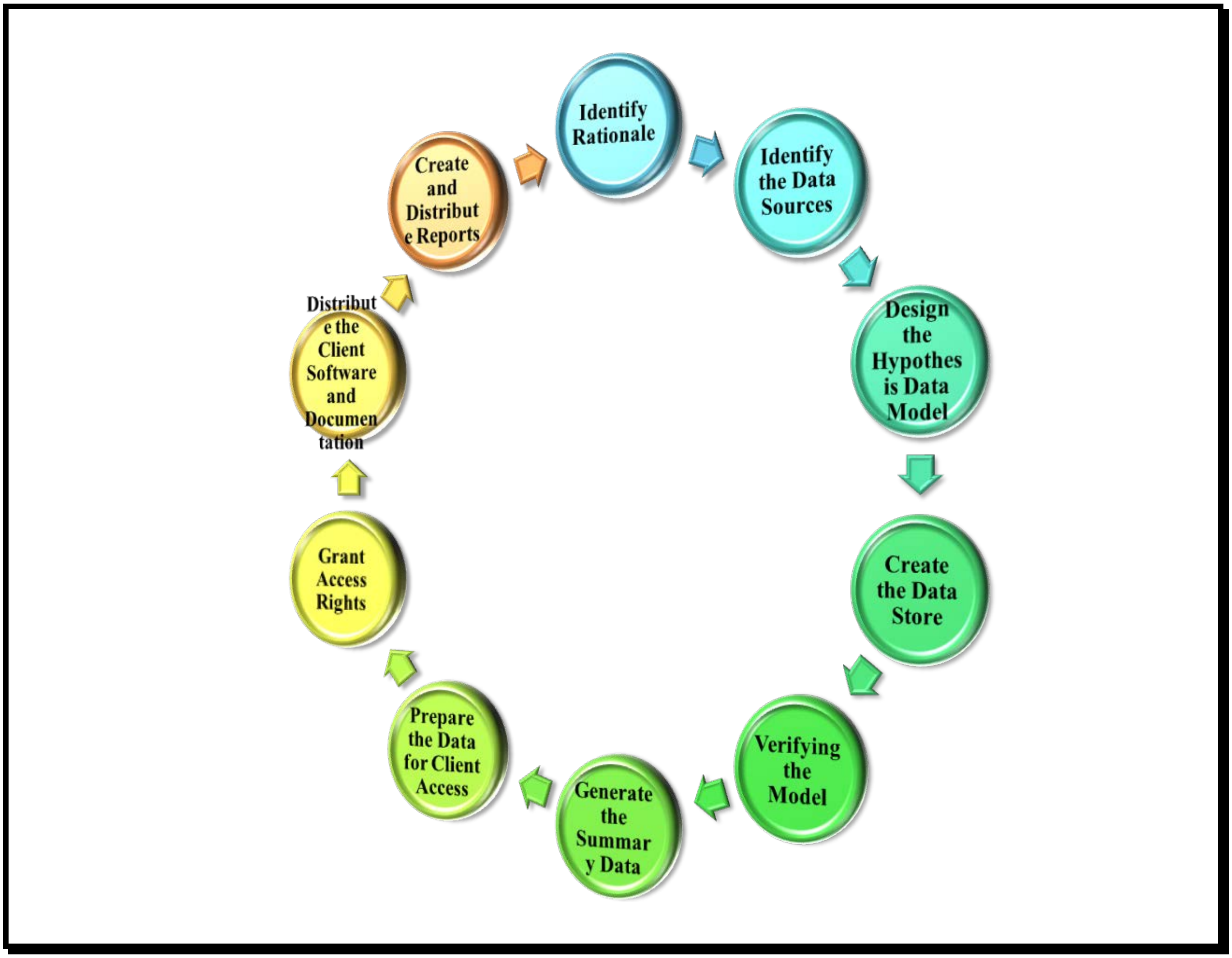

Fig. 1 A Cycle of BI Formulation

H.Grant Access Rights

Users must have database access rights granted to them so that they can view and manipulate the data. BI should be extremely protected based upon access rights and ensuring right information for right person through right authority procedures.

I. Distribute the Client Software and Documentation

After the data store is ready for client access, you can distribute the software and provide documentation to your end users.

J. Create and Distribute Reports

Report developers can develop reports and share them with the user community. For reports created in Discoverer Plus, you can create a dashboard where reports can be published.

III. BI STRUCTURING

Structural equation modeling, SEM, has gained popularity across many disciplines in the past two decades due perhaps to its generally and flexibility. As a statistical modeling tool, its development and expansion are rapid and ongoing.

With advances in estimation techniques, basic models, such as measurement models, path models, and their integration into a general covariance structure SEM analysis framework have been expanded to include, but are by no means limited to, the modeling of mean structures, interaction or nonlinear relations, and multilevel problems. The purpose of this study was to identify the foundations of SEM modeling with the basic covariance structure models could be applied on BI structuring.

A. Structural Equation Modeling

Structural equation modeling is a general term that has been used to describe a large number of statistical models used to evaluate the validity of substantive theories with empirical data. Statistically, it represents an extension of general linear modeling (GLM) procedures, such as the ANOVA and multiple regression analysis. One of the primary advantages of SEM (vs. other applications of GLM) is that it can be used to study the relationships among latent constructs that are indicated by multiple measures. It is also applicable to both experimental and non-experimental data, as well as cross-sectional and longitudinal data. SEM takes a confirmatory (hypothesis testing) approach to the multivariate analysis of a structural theory, one that stipulates causal relations among multiple variables. The causal pattern of inter-variable relations within the theory is specified a priori. The goal is to determine whether a hypothesized theoretical model is consistent with the data collected to reflect this theory. The consistency is evaluated through model-data fit, which indicates the extent to which the postulated network of relations among variables is plausible[10].

SEM is a large sample technique (usually $N > 200$; e.g., [11]) and the sample size required is somewhat dependent on model complexity, the estimation method used, and the distributional characteristics of observed variables [11]. SEM has a number of synonyms and special cases in the literature including path analysis, causal modeling, and covariance structure analysis. In simple terms, SEM involves the evaluation of two models: a measurement model and a path model.

B. Path Model

Path analysis is an extension of multiple regression in that it involves various multiple regression models or equations that are estimated simultaneously. This provides a more effective and direct way of modeling mediation, indirect effects, and other complex relationship among variables. Path analysis can be considered a special case of SEM in which structural relations among observed (vs. latent) variables are modeled.

Structural relations are hypotheses about directional influences or causal relations of multiple variables (e.g., how independent variables affect dependent variables). Hence, path analysis (or the more generalized SEM) is sometimes referred to as causal modeling. Because analyzing interrelations among variables is a major part of SEM and these interrelations are hypothesized to generate specific observed covariance (or correlation) patterns among the variables, SEM is also sometimes called covariance structure analysis.

In SEM, a variable can serve both as a source variable (called an exogenous variable, which is analogous to an independent variable) and a result variable (called an endogenous variable, which is analogous to a dependent variable) in a chain of causal hypotheses. This kind of variable is often called a mediator. As an example, suppose that family environment has a direct impact on learning motivation which, in turn, is hypothesized to affect achievement. In this case motivation is a mediator between family environment and achievement; it is the source variable for achievement and the result variable for family environment. Furthermore, feedback loops among variables (e.g., achievement can in turn affect family environment in the example) are permissible in SEM, as are reciprocal effects (e.g., learning motivation and achievement affect each other).

In path analyses, observed variables are treated as if they are measured without error, which is an assumption that does not likely hold in most social and behavioral sciences. When observed variables contain error, estimates of path coefficients may be biased in unpredictable ways, especially for complex models [10, 12]. Estimates of reliability for the measured variables, if available, can be incorporated into the model to fix their error variances (e.g., squared standard error of measurement via classical test theory). Alternatively, if multiple observed variables that are supposed to measure the same latent constructs are available, then a measurement model can be used to separate the common variances of the observed variables from their error variances thus correcting the coefficients in the model for unreliability.

C. Measurement Model

The measurement of latent variables originated from psychometric theories. Unobserved latent variables cannot be measured directly but are indicated or inferred by responses to a number of observable variables (indicators). Latent constructs such as intelligence or reading ability are often gauged by responses to a battery of items that are designed to tap those constructs. Responses of a study participant to those items are supposed to reflect where the participant stands on the latent variable. Statistical techniques, such as factor analysis, exploratory or confirmatory, have been widely used to examine the number of latent constructs underlying the observed responses and to evaluate the adequacy of individual items or variables as indicators for the latent constructs they are supposed to measure.

The measurement model in SEM is evaluated through confirmatory factor analysis (CFA). CFA differs from exploratory factor analysis (EFA) in that factor structures are hypothesized a priori and verified empirically rather than derived from the data. EFA often allows all indicators to load on all factors and does not permit correlated residuals. Solutions for different number of factors are often examined in EFA and the most sensible solution is interpreted. In contrast, the number of factors in CFA is assumed to be known. In SEM, these factors correspond to the latent constructs represented in the model. CFA allows an indicator to load on multiple factors (if it is believed to measure multiple latent constructs). It also allows residuals or errors to correlate (if these indicators are believed to have common causes other than the latent factors included in the model). Once the measurement model has been specified, structural relations of the latent factors are then modeled essentially the same way as they are in path models. The combination of CFA models with structural path models on the latent constructs represents the general SEM framework in analyzing covariance structures.

D. BI Structure Procedure

In general, every SEM analysis goes through the steps of model specification, data collection, model estimation, model evaluation, and (possibly) model modification. Issues pertaining to each of these steps are discussed below.

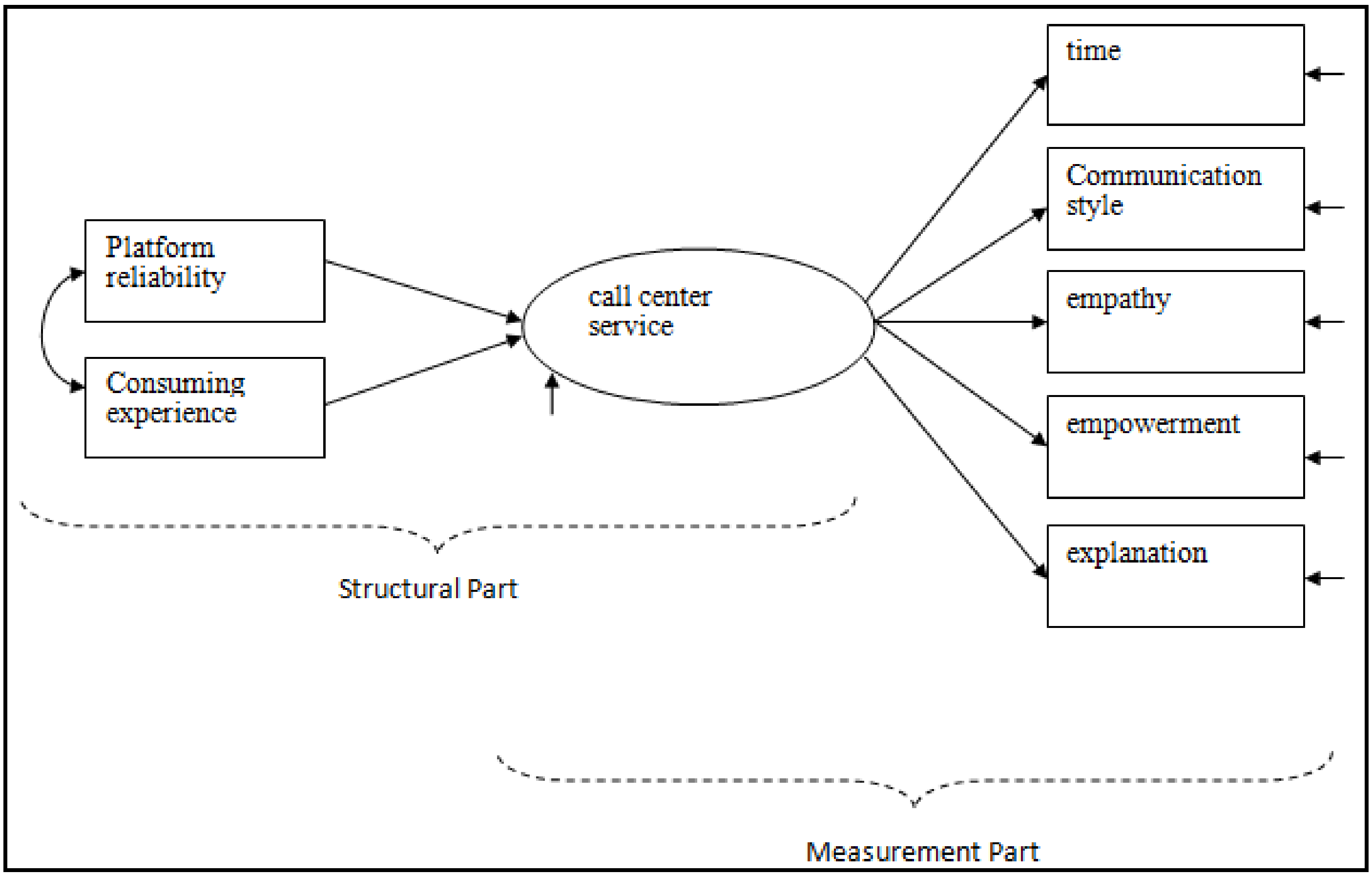

Fig. 2 Conceptual model of a call center service

E. BI specification

A sound model is theory based. Theory is based on findings in the BI literature, knowledge in the field, or one's educated guesses, from which causes and effects among variables within the BI statements are specified. Models are often easily conceptualized and communicated in graphical forms. In these graphical forms, a directional arrow (→) is universally used to indicate a hypothesized causal direction. The variables to which arrows are pointing are commonly termed endogenous variables (or dependent variables) and the variables having no arrows pointing to them are called exogenous variables (or independent variables). Unexplained covariances among variables are indicated by curved arrows. Observed variables are commonly enclosed in rectangular boxes and latent constructs are enclosed in circular or elliptical shapes.

For example, suppose a group of researchers have developed a new measure to assess call center services of an on-line information platform and would like to find out (a) whether the service scores measure a common construct called call center service and (b) whether platform reliability has an influence on call center service when consuming experience (measured in month) differences are controlled for. The call center service scores available are: time, communication style, empathy, empowerment, and explanation.

These service scores (indicators) are hypothesized to indicate the strength of call center service, with higher scores signaling stronger service. Figure 1 presents the conceptual model.

The model in Figure 2 suggests that the five service scores on the right are supposedly results of latent call center service (enclosed by an oval) and that the two exogenous observed variables on the left (Platform reliability and Consuming experience by rectangles) are predictors of call center service. The two predictors (connected by) are allowed to be correlated but their relationship is not explained in the model. The latent “call center service” variable and the five observed skill scores (enclosed by rectangles) are endogenous in this example. The residual of the latent endogenous variable (residuals of structural equations are also called disturbances) and the residuals (or errors) of the service variables are considered exogenous because their variances and interrelationships are unexplained in the model. The residuals are indicated by arrows without sources in Figure 2. The effects of Platform reliability and Consuming experience in month on the five service scores can also be perceived to be mediated by the latent variable (call center service). This model is an example of a multiple-indicator multiple-cause model (or MIMIC for short, a special case of general SEM model) in which the service scores are the indicators and Consuming experience as well as Platform reliability are the causes for the latent variable.

Due to the flexibility in model specification, a variety of models can be conceived. However, not all specified models can be identified and estimated. Just like solving equations in algebra where there cannot be more unknowns than knowns, a basic principle of identification is that a model cannot have a larger number of unknown parameters to be estimated than the number of unique pieces of information provided by the data (variances and covariances of observed variables for covariance structure models in which mean structures are not analyzed). Because the scale of a latent variable is arbitrary, another basic principle of identification is that all latent variables must be scaled so that their values can be interpreted. These two principles are necessary for identification but they are not sufficient. The issue of model identification is complex. Fortunately, there are some established rules that can help researchers decide if a particular model of interest is identified or not [12, 13].

When a model is identified for the BI, every model parameter can be uniquely estimated. A model is said to be over-identified if it contains fewer parameters to be estimated than the number of variances and covariances, just-identified when it contains the same number of parameters as the number of variances and covariances, and under-identified if the number of variances and covariances is less than the number of parameters. Parameter estimates of an over-identified model are unique given a certain estimation criterion (e.g., maximum likelihood). All just-identified models fit the data perfectly and have a unique set of parameter estimates. However, a perfect model-data fit is not necessarily desirable in SEM. First, sample data contain random error and a perfect-fitting model may be fitting sampling errors. Second, because conceptually very different just-identified models produce the same perfect empirical fit, the models cannot be evaluated or compared by conventional means (model fit indices discussed below). When a model cannot be identified, either some model parameters cannot be estimated or numerous sets of parameter values can produce the same level of model fit (as in under-identified models). In any event, results of such models are not interpretable and the models require re-specification.

F. Data Characteristics

Like conventional statistical techniques, score reliability and validity should be considered in selecting measurement instruments for the constructs of interest and sample size needs to be determined preferably based on power considerations. The sample size required to provide unbiased parameter estimates and accurate model fit information for SEM models depends on model characteristics, such as model size as well as score characteristics of measured variables, such as score scale and distribution. For example, larger models require larger samples to provide stable parameter estimates, and larger samples are required for categorical or non-normally distributed variables than for continuous or normally distributed variables. Therefore, data collection should come, if possible, after models of interest are specified so that sample size can be determined a priori. Information about variable distributions can be obtained based on a pilot study or one’s educated guess.

SEM is a large sample technique. That is, model estimation and statistical inference or hypothesis testing regarding the specified model and individual parameters are appropriate only if sample size is not too small for the estimation method chosen. A general rule of thumb is that the minimum sample size should be no less than 200 (preferably no less than 400 especially when observed variables are not multivariate normally distributed) or 5–20 times the number of parameters to be estimated, whichever is larger [11]. Larger models often contain larger number of model parameters and hence demand larger sample sizes. Sample size for SEM analysis can also be determined based on a priori power considerations. There are different approaches to power estimation in SEM [14-18].

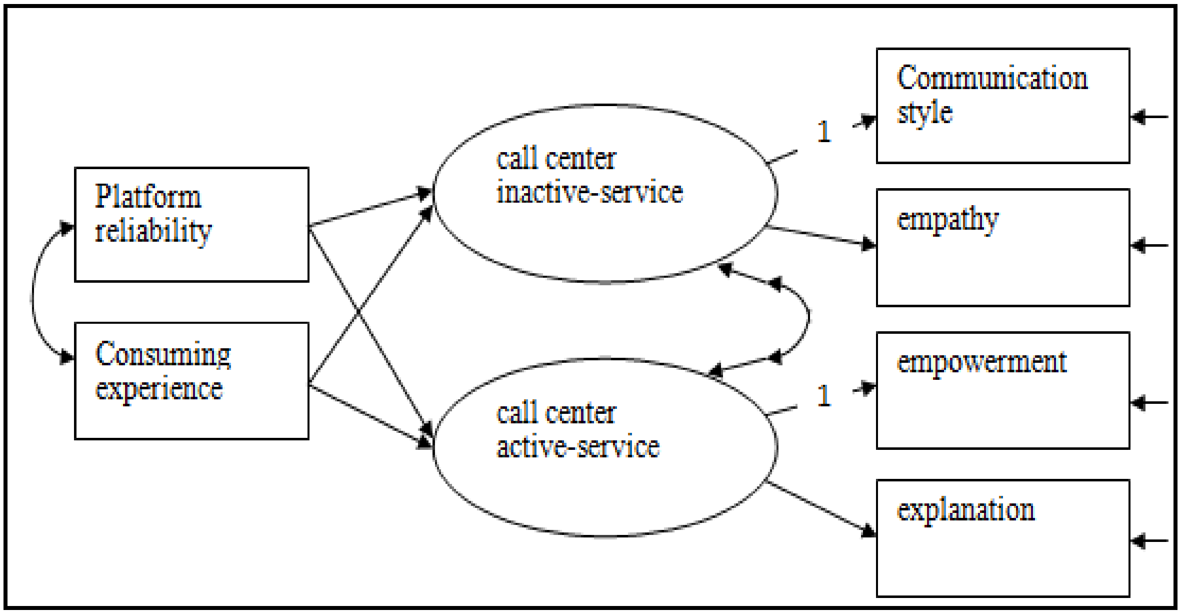

Fig. 3 Conceptual model of inactive-service and active service

G. Structure Estimation

A properly specified structural equation model often has some fixed parameters and some free parameters to be estimated from the data. As an illustration, Figure 3 shows the diagram of a conceptual model that predicts inactive-service (IAS) and active-service (AS) latent call center service from observed scores from two consumer scales, Platform reliability (PR) and Consuming experience (CE). The latent inactive service variable is indicated by communication style (CS) and empathy scores (ES). The latent active-service variable is indicated by empowerment (EM) and explanation (EX). The visible paths denoted by directional arrows (from PR and CE to IAS and AS, from IAS to CS and ES, and from AS to EM and EX) and curved arrows (between PR and CE, and between residuals of IAS and AS) in the diagram are free parameters of the model to be estimated, as are residual variances of endogenous variables (IAS, AS, CS, ES, EM, and EX) and variances of exogenous variables (PR and CE). All other possible paths that are not shown (e.g., direct paths from PR or CE to CS, ES, EM, or EX) are fixed to zero and will not be estimated. As mentioned above, the scale of a latent variable is arbitrary and has to be set. The scale of a latent variable can be standardized by fixing its variance to 1. Alternatively, a latent variable can take the scale of one of its indicator variables by fixing the factor loading (the value of the path from a latent variable to an indicator) of one indicator to 1. In this example, the loading of CS on IAS and the loading of EM on AS are fixed to 1 (i.e., they become fixed parameters). That is, when the parameter value of a visible path is fixed to a constant, the parameter is not estimated from the data.

Free parameters are estimated through iterative procedures to minimize a certain discrepancy or fit function between the observed covariance matrix (data) and the model-implied covariance matrix (model). Definitions of the discrepancy function depend on specific methods used to estimate the model parameters. A commonly used normal theory discrepancy function is derived from the maximum likelihood method. This estimation method assumes that the observed variables are multivariate normally distributed or there is no excessive kurtosis (i.e., same kurtosis as the normal distribution) of the variables [12].

The estimation of a model may fail to converge or the solutions provided may be improper. In the former case, SEM software programs generally stop the estimation process and issue an error message or warning. In the latter, parameter estimates are provided but they are not interpretable because some estimates are out of range (e.g., correlation greater than 1, negative variance). These problems may result if a model is ill specified (e.g., the model is not identified), the data are problematic (e.g., sample size is too small, variables are highly correlated, etc.), or both. Multicollinearity occurs when some variables are linearly dependent or strongly correlated (e.g., bivariate correlation > .85). It causes similar estimation problems in SEM as in multiple regression. Methods for detecting and solving multicollinearity problems established for multiple regression can also be applied in SEM.

Table 2. Sample Correlation, Mean, and Standard Deviation for the Model of Fig.1

	Variables						
Variables	PR	CE	T	CS	ES	EM	EX
Platform reliability (PR)	1						
Consuming experience (CE)	.357	1					
Time (T)	.382	.439	1				
Communication style (CS)	.510	.405	.588	1			
Empathy Score (ES)	.439	.447	.512	0.604	1		
Empowerment (EM)	.513	.475	.591	.560	.606	1	
Explanation (EX)	.372	.328	.564	.531	.443	.561	1
Mean	50.340	.440	.666	.730	.545	.625	.624
SD	6.706	1.023	1.297	.855	.952	.933	1.196

H. Structure Evaluation

Once model parameters have been estimated, one would like to make a dichotomous decision, either to retain or reject the hypothesized model. This is essentially a statistical hypothesis-testing problem, with the null hypothesis being that the model under consideration fits the data. The overall model goodness of fit is reflected by the magnitude of discrepancy between the sample covariance matrix and the covariance matrix implied by the model with the parameter estimates (also referred to as the minimum of the fit function or Fmin). Most measures of overall model goodness of fit are functionally related to Fmin. The model test statistic (N – 1)Fmin, where N is the sample size, has a chi-square distribution (i.e., it is a chi-square test) when the model is correctly specified and can be used to test the null hypothesis that the model fits the data. Unfortunately, this test statistic has been found to be extremely sensitive to sample size. That is, the model may fit the data reasonably well but the chi-square test may reject the model because of large sample size.

In reaction to this sample size sensitivity problem, a variety of alternative goodness-of-fit indices have been developed to supplement the chi-square statistic. All of these alternative indices attempt to adjust for the effect of sample size, and many of them also take into account model degrees of freedom, which is a proxy for model size. Two classes of alternative fit indices, incremental and absolute, have been identified [10]. Incremental fit indices measure the increase in fit relative to a baseline model (often one in which all observed variables are uncorrelated). Examples of incremental fit indices include normed fit index, Tucker-Lewis index, relative noncentrality index, and comparative fit index [10, 15]. Higher values of incremental fit indices indicate larger improvement over the baseline model in fit. Values in the .90s (or more recently≥.95) are generally accepted as indications of good fit.

In contrast, absolute fit indices measure the extent to which the specified model of interest reproduces the sample covariance matrix. Examples of absolute fit indices include goodness-of-fit index (GFI) and adjusted GFI (AGFI), standardized root mean square residual, and the RMSEA. Higher values of GFI and AGFI as well as lower values of SRMR and RMSEA indicate better model-data fit[10, 15].

SEM software programs routinely report a handful of goodness-of-fit indices. Some of these indices work better than others under certain conditions. It is generally recommended that multiple indices be considered simultaneously when overall model fit is evaluated[10].

Because some solutions may be improper, it is prudent for researchers to examine individual parameter estimates as well as their estimated standard errors. Unreasonable magnitude (e.g., correlation>1) or direction (e.g., negative variance) of parameter estimates or large standard error estimates (relative to others that are on the same scale) are some indications of possible improper solutions.

If a model fits the data well and the estimation solution is deemed proper, individual parameter estimates can be interpreted and examined for statistical significance (whether they are significantly different from zero). The test of individual parameter estimates for statistical significance is based on the ratio of the parameter estimate to its standard error estimate (often called z-value or t-value). As a rough reference, absolute value of this ratio greater than 1.96 may be considered statistically significant at the .05 level. Although the test is proper for unstandardized parameter estimates, standardized estimates are often reported for ease of interpretation. In growth models and multiple-sample analyses in which different variances over time or across samples may be of theoretical interest, unstandardized estimates are preferred.

As an example, Table 2 presents the simple descriptive statistics of the variables for the math ability example (Figure 1), and Table 3 provides the parameter estimates (standardized and unstandardized) and their standard error estimates. This model fit the sample data reasonably well as indicated by the selected overall goodness-of-fit statistics: $\chi 2$ = 21.21, p = .069, RMSEA = .056 (<.06), CFI = .99 (>.95), SRMR = .032 (<.08). The model solution is considered proper because there are no out-of-range parameter estimates and standard error estimates are of similar magnitude (see Table 2). All parameter estimates are considered large (not likely zero) because the ratios formed by unstandardized parameter estimates to their standard errors (i.e., z-values or t-values) are greater than |2| [11].

Table 3. Parameter and Standard Error Estimates for the Model of Fig.1.

Model Parameters	Standardized Estimate	Un-standardized Estimate	Standard Error
Loadings / effects on call center service			
T	.74	1.00	
CS	.77	.68	.07
ES	.74	.73	.08
EM	.80	.77	.08
EX	.68	.84	.10
PR	.46	.07	.01
CE	.40	.38	.07
Residual variances			
T	.45	.75	.10
CS	.41	.30	.04
ES	.46	.41	.05
EM	.36	.32	.05
EX	.54	.78	.09
Call center service	.50	.46	.09
Covariance			
PR and CE	.36	2.45	.56

Standardized factor loadings in measurement models should fall between 0 and 1 with higher values suggesting better indications of the observed variables for the latent variable. All standardized loadings in this example are in the neighborhood of .7, showing that they are satisfactory indicators for the latent construct of math ability. Coefficients for the structural paths are interpreted in the same way as regression coefficients. The standardized coefficient value of .46 for the path from consuming experience to call center service suggests that as user's experience grow by one standard deviation of experience in months (about 6.7 months), their call center service score is expected to increase by .46 standard deviation holding PR constant. The standardized value of .40 for the path from PR to call center service score reveals that for every standard deviation increase in PR, call center service score is expected to increase by .40 standard deviation, holding consuming experience constant. The standardized residual variance of .50 for the latent call center service variable indicates that approximately 50% of variance in service score is unexplained by PR and CE. Similarly, standardized residual or error variances of the math indicator variables are taken as the percentages of their variances unexplained by the latent variable.

I. Modification, Alternative and Equivalent Structure

When the hypothesized model is rejected based on goodness of fit statistics, SEM researchers are often interested in finding an alternative model that fits the data. Post hoc modifications (or model trimming) of the model are often aided by modification indices, sometimes in conjunction with the expected parameter change statistics. Modification index estimates the magnitude of decrease in model chi-square (for 1 degrees of freedom) whereas expected parameter change approximates the expected size of change in the parameter estimate when a certain fixed parameter is freely estimated. A large modification index (>3.84) suggests that a large improvement in model fit as measured by chi-square can be expected if a certain fixed parameter is freed. The decision of freeing a fixed parameter is less likely affected by chance if it is based on a large modification index as well as a large expected parameter change value.

As an illustration, Table 3 shows the simple descriptive statistics of the variables for the model of Figure 2, and Table 4 provides the parameter estimates (standardized and unstandardized) and their standard error estimates. Had one restricted the residuals of the latent READ and MATH variables to be uncorrelated, the model would not fit the sample data well as suggested by some of the overall model fit indices: χ 26 =45.30, p<.01, RMSEA=.17 (>.10), SRMR= .078 (acceptable because it is < .08). The solution was also improper because there was a negative error variance estimate. The modification index for the covariance between the residuals of READ and MATH was 33.03 with unstandardized expected parameter change of 29.44 (standardized expected change = .20). There were other large modification indices. However, freeing the residual covariance between READ and MATH was deemed most justifiable because the relationship between these two latent variables was not likely fully explained by the two intelligence subtests (VC and PO). The modified model appeared to fit the data quite well (χ 25 = 8.63, p =.12, RMSEA = .057, SRMR =.017). The actual chisquare change from 45.30 to 8.63 (i.e., 36.67) was slightly different from the estimated change (33.03), as was the actual parameter change (31.05 vs. 29.44; standardized value= .21 vs. .20). The differences between the actual and estimated changes are slight in this illustration because only one parameter was changed. Because parameter estimates are not independent of each other, the actual and expected changes may be very different if multiple parameters are changed simultaneously, or the order of change may matter if multiple parameters are changed one at a time. In other words, different final models can potentially result when the same initial model is modified by different analysts.

As a result, researchers are warned against making a large number of changes and against making changes that are not supported by strong substantive theories [10, 17, 18]. Changes made based on modification indices may not lead to the “true” model in a large variety of realistic situations [10, 17, 18]. The likelihood of success of post hoc modification depends on several conditions: It is higher if the initial model is close to the “true” model, the search continues even when a statistically plausible model is obtained, the search is restricted to paths that are theoretically justifiable, and the sample size is large [17]. Unfortunately, whether the initially hypothesized model is close to the “true” model is never known in practice. Therefore, one can never be certain that the modified model is closer to the “true” model.

Moreover, post hoc modification changes the confirmatory approach of SEM. Instead of confirming or disconfirming a theoretical model, modification searches can easily turn modeling into an exploratory expedition. The model that results from such searches often capitalizes on chance idiosyncrasies of sample data and may not generalize to other samples. Hence, not only is it important to explicitly account for the specifications made post hoc, but it is also crucial to cross-validate the final model with independent samples.

Rather than data-driven post hoc modifications, it is often more defensible to consider multiple alternative models a priori. That is, multiple models (e.g., based on competing theories or different sides of an argument) should be specified prior to model fitting and the best fitting model is selected among the alternatives. As models that are just-identified will fit the data perfectly regardless of the particular specifications, different just-identified models (sub-models or the entire model) detailed for the same set of variables are considered equivalent. Equivalent models may be very different in implications but produce identical model-data fit. For instance, predicting verbal ability from quantitative ability may be equivalent to predicting quantitative ability from verbal ability or to equal strength of reciprocal effects between verbal and quantitative ability. In other words, the direction of causal hypotheses cannot be ruled out (or determined) on empirical grounds using cross-sectional data but on theoretical foundations, experimental control, or time precedence if longitudinal data are available.

Researchers are encouraged to consider different models that may be empirically equivalent to their selected final model(s) before they make any substantial claims.

J. Causal Relations

Although SEM allows the testing of causal hypotheses, a well fitting SEM model does not and cannot prove causal relations without satisfying the necessary conditions for causal inference, partly because of the problems of equivalent models discussed above. The conditions necessary to establish causal relations include time precedence and robust relationship in the presence or absence of other variables. A selected well-fitting model in SEM is like a retained null hypothesis in conventional hypothesis testing. It remains plausible among perhaps many other models that are not tested but may produce the same or better level of fit. SEM users are cautioned not to make unwarranted causal claims. Replications of findings with independent samples are essential especially if the models are obtained based on post hoc modifications. Moreover, if the models are intended to be used in predicting future behaviors, their utility should be evaluated in that context.

IV. CONCLUSION

The results of this study provide an concrete evidence for the feasibility of integrating cloud computing into senior high-school learning. Following the strict learning goal and principles of technology education, the content selecting and learning experience could be ideally organized for fitting into original curriculum.

As the finding shown, there are four level knowledge of cloud computing technology provided for learning. Those four levels of knowledge are fact, concept, procedure, and meta-analysis. In this study, the fact knowledge mostly comes from definition.

Based upon definitions of cloud computing technology, this emerging technology become reality and could be further discussed and explained. The learning goals, knowledge structure, learning experience then could be identified to organize the subject matter. The content of cloud computing were identified and verified with the curriculum standard by this study.

The purpose of this study was to identify the content structure and learning experience for integrating cloud computing into high-school learning offered by the High-Scope Project. This study provided a response of coping emerging technology by integrating cloud computing into formal education.

By introducing learns with following topics, a core technology education course of integrating cloud computing are designed. Those topic are:

- Technology Development
- Technology World
- Creative Design Production

An advanced communication technology course is also designed. The major contents are:

- Electronic Communication
- Information Communication
- Communication Ethics
- Communication Industry
- Project of Design & Production

It is concluded that cloud computing could be integrated into high-school technology education as an emerging technology. The in-service teacher education might be used for promoting this new course[19-21]

REFERENCES

[1] A. D. and H. D., "Data Visualization in Business Intelligence," in *Proceedings of the 11th WSEAS Int. Conf. on Mathematics and Computers in Business and Economics MCBE'10*, ed: WSEAS, 2011, pp. 164-167.

[2] L. M. Chen, L. H. Kuo, and H. J. Yang, "Applying Computerized Digitizing Technique to Explore the POP Album Cover Historical Reflections," *INTERNATIONAL JOURNAL OF COMMUNICATIONS*, vol. 6, pp. 109-119, 2012.

[3] L. H. Kuo, H. M. Wei, L. M. Chen, M. C. Wang, M. K. Ho, and H. J. Yang, "An Evaluation Model of Integrating Emerging Technology into Formal Curriculum," *INTERNATIONAL JOURNAL OF EDUCATION AND INFORMATION TECHNOLOGIES*, vol. 6, pp. 250-259, 2012.

[4] L. H. Kuo, H. M. Wei, W. C. Hu, and H. J. Yang, "Applying Innovation Theory in Observing Emerging Technology Acceptance," *International Journal of Systems Applications, Engineering & Development*, vol. 7, p. 56~65, 2013.

[5] L. H. Kuo, J. C. Yu, H. H. Yang, W. C. Hu, and H. J. Yang, "A Study of Creating Technology Education Course for Cloud Computing," *INTERNATIONAL JOURNAL OF Communications,*, vol. 6, pp. 98-108, 2012.

[6] L.-H. Kuo, H.-J. Yang, L. Lin, and H.-C. Lin, "Identifying a General Structure of Teachers' On-line In-service Learning," in *10th WSEAS International Conference on Edication and Educational Technology (EDU'11)*, Penang, Malaysia, 2011, pp. 87-92.

[7] L.-H. Kuo, H.-J. Yang, and Y.-W. L. Lin, "Overcoming the imbalance in the supply and demand of professionals in the marine industry: Professional development of marine education in Taiwan," *African Journal of Business Management*, vol. 6, p. 9202~9209, 2012.

[8] T. D. and M. M., "A Business Intelligence Virtual Competency Community of Practice Proposal, ," in *Proceedings of IADIS International Conferences: Web Based Communities 2011, Collaborative Technologies 2011 and InternetApplications and Research 2011*, ed Rome, Italy: IADIS, 2011, pp. 249-252.

[9] O. H. D. and B. P.-D., "Business Intelligence and Information Systems:Enhancing Student Knowledge in Database Courses," *Review of Business Information Systems*, vol. 16, pp. 1-14, First Quarter 2012 2012.

[10] B. M. Byrne, *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates, 1998.

[11] R. B. Kline, *Principles and practice of structural equation modeling*, 2nd ed. New York: Guilford Press, 2005.
[12] P. M. Bentler, "Comparative fit indexes in structural models," *Psychological Bulletin*, vol. 107, pp. 238-246, 1990.
[13] J. Pearl, *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press, 2000.
[14] L. K. Muth´en and B. O. Muth´en, "How to use a Monte Carlo study to decide on sample size and determine power," *Structural Equation Modeling*, vol. 4, pp. 599-620, 2002.
[15] J. L. Arbuckle, "Amos 5," ed. Chicago, IL: Smallwaters, 2003.
[16] M. C. Neale, S. M. Boker, G. Xie, and H. H. Maes, *Mx: Statistical modeling*, 6 ed. Richmond, VA: Department of Psychiatry, Virginia Commonwealth University., 2003.
[17] H.-J. Yang and H.-H. Yang, "Developing IT Knowledge Transfer Model by Using SEM," *WSEAS Transactions on Mathematics,* vol. 5, pp. 174-181, 2006.
[18] H. H. Yang, J. C. Yu, and H. J. Yang, "A Study of Verifying Knowledge Reuse Path Using SEM," in *the European Computing Conference 2007 Proceedings, September 25-27, 2007*, E. C. Conference, Ed., ed Vouliagmeni Beach, Athens, Greece,: WSEAS, 2007.
[19] L. H. Kuo, H. M. Wei, L. M. Chen, M. C. Wang, M. K. Ho, and H. J. Yang, "An Evaluation Model of Integrating Emerging Technology into Formal Curriculum," *INTERNATIONAL JOURNAL OF EDUCATION AND INFORMATION TECHNOLOGIES,* vol. 6, p. 10, 2012.
[20] L. H. Kuo, J. C. Yu, H. H. Yang, W. C. Hu, and H. J. Yang, "A Study of Creating Technology Education Course for Cloud Computing," *INTERNATIONAL JOURNAL OF Communications,* vol. 6, p. 11, 2012.
[21] L. H. Kuo, J. C. Yu, H. J. Yang, and L. Lin, "ICT Learning Groups among Teachers In-service Education Participants," *INTERNATIONAL JOURNAL OF EDUCATION AND INFORMATION TECHNOLOGIES,* vol. 6, p. 9, 2012.

Lung-Hsing Kuo received his Master (M.E.) in Education (1990~1993) and Ph.D. in Education from (1993~1997) National Kaohsiung Normal University. He is the director of the center for teacher career and professional development in National Kaohsiung Normal University. His research interests include social Science Research Methodology, Continuing Education, Human and social, Youth Study, Emotion development and management, Counseling and Education Issues.

Hsieh-Hua Yang received her Master (M.S.) in Health Promotion and Health Education from National Taiwan Normal University and Ph.D. in Health Policy and Management from National Taiwan University. She is a professor in Health Care Administration at Oriental Institute of Technology. Her research focuses are health behavior, social network and organizational behavior.

Hsueh-Chih Lin received a Master (M.S.) in Industrial Technology Education from National Kaohsiung Normal University on the year of 2006. He received a Ph.D. in the department Industrial Technology Education at National Kaohsiung Normal University on the year of 2013. His research is focus on Design and technology, technology development and technology education.

Miao-Kuei Ho received her Master (M.S.) in Industrial Technology Education (2003~2005) from National Kaohsiung Normal University. She is pursuing her Ph.D. in Industrial Technology Education from National Kaohsiung Normal University now. She has been teaching elementary school since 1990. As a senior teacher, she has had ten years field research experience with NSC projects. Her research interests include technology education, educational technology, and learning theory. She also has experts in technology process skills.

Hung-Jen Yang obtained a Master (M.S.) in Technology Education from University of North Dakota and a Ph.D. in Industrial Technology Education from Iowa State University. He is currently conducting research on knowledge transfer, and knowledge reuse via information technology. His research has appeared in a variety of journals including those published by the WSEAS.