Structural Equation Modeling: An Introduction

Prof. Dr. Olaf Rank

Structural Equations Modeling Defined

Structural Equations Modeling (SEM)

Multivariate technique combining aspects of factor analysis and multiple regression that enables the researcher to simultaneously examine a series of interdependence relationships among the *measured variables* and *latent constructs* as well as between several latent constructs.

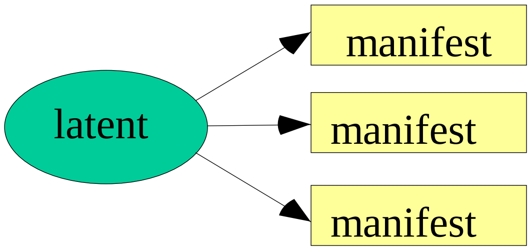
Constructs are unobservable or latent factors that are represented by multiple measured variables.

Distinguishing Features of SEM

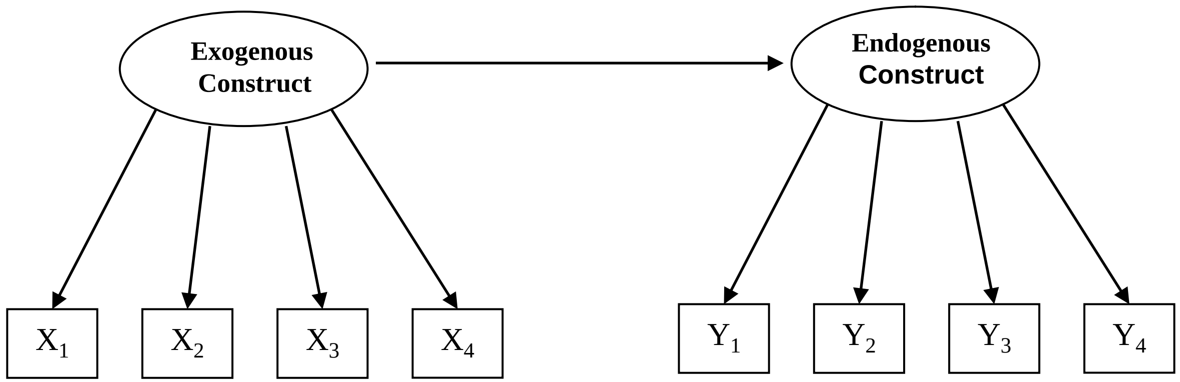
- Estimation of multiple and interrelated relationships.
- Represents unobserved (latent) concepts and corrects for measurement error.
- Defines a model to explain an entire set of relationships.

Latent Constructs and Abbreviations

- Exogenous constructs are the latent, multi-item equivalent of independent variables.
- Endogenous constructs are the latent, multi-item equivalent to dependent variables.
- These constructs are theoretically determined by factors within the model.
- Multiple measured variables (x and y) represent the exogenous and endogenous constructs.

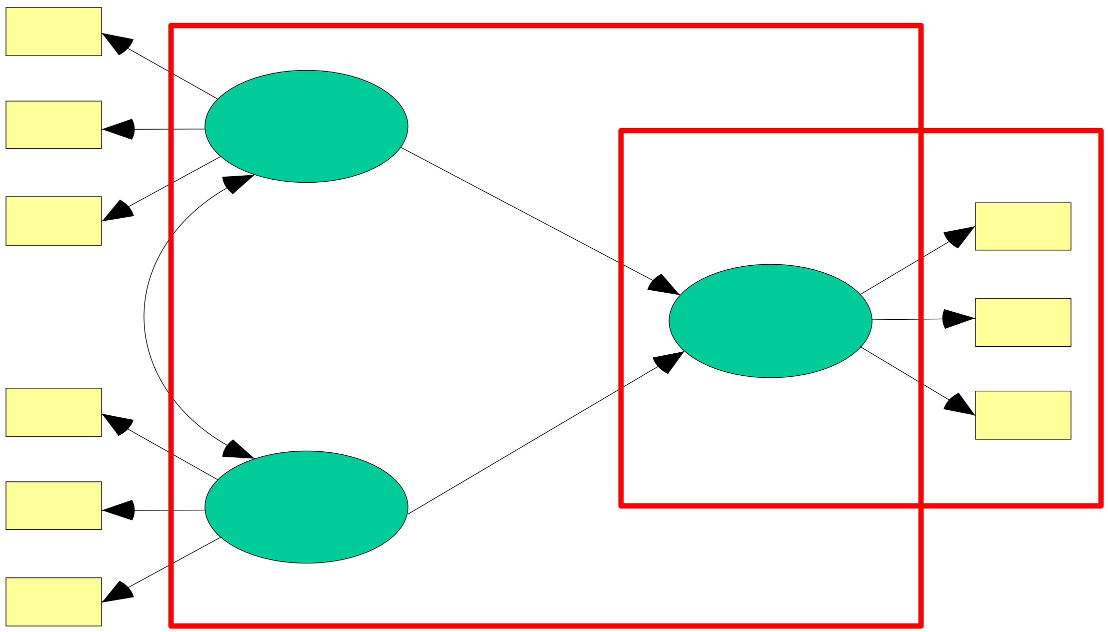


Two Latent Constructs and the Measured Variables that Represent Them



- Loadings represent the relationships from constructs to variables as in factor analysis.
- Path estimates represent the relationships between constructs as does B in regression analysis.

Measurement and Structural Models

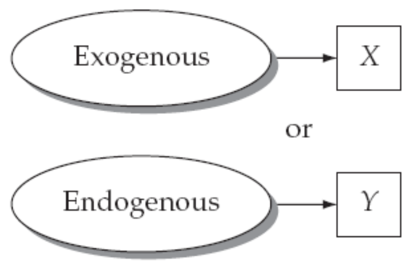


Measurement (outer) model: Relationships between and among observed and latent variables (items and factors)

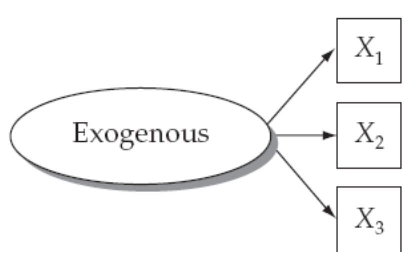
Structural (inner) model: Relationships between latent variables

Distinguishing the Types of Relationships Involved in SEM

A. Relationship between a construct and a measured variable

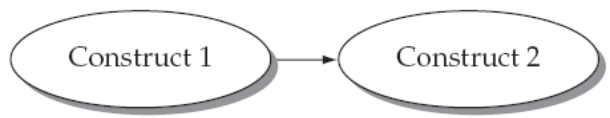


B. Relationships between a construct and multiple measured variables

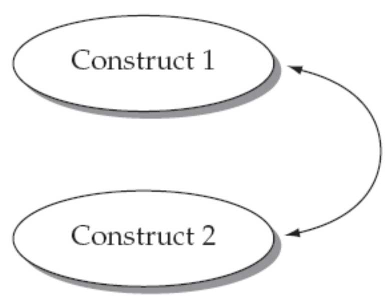


Distinguishing the Types of Relationships Involved in SEM

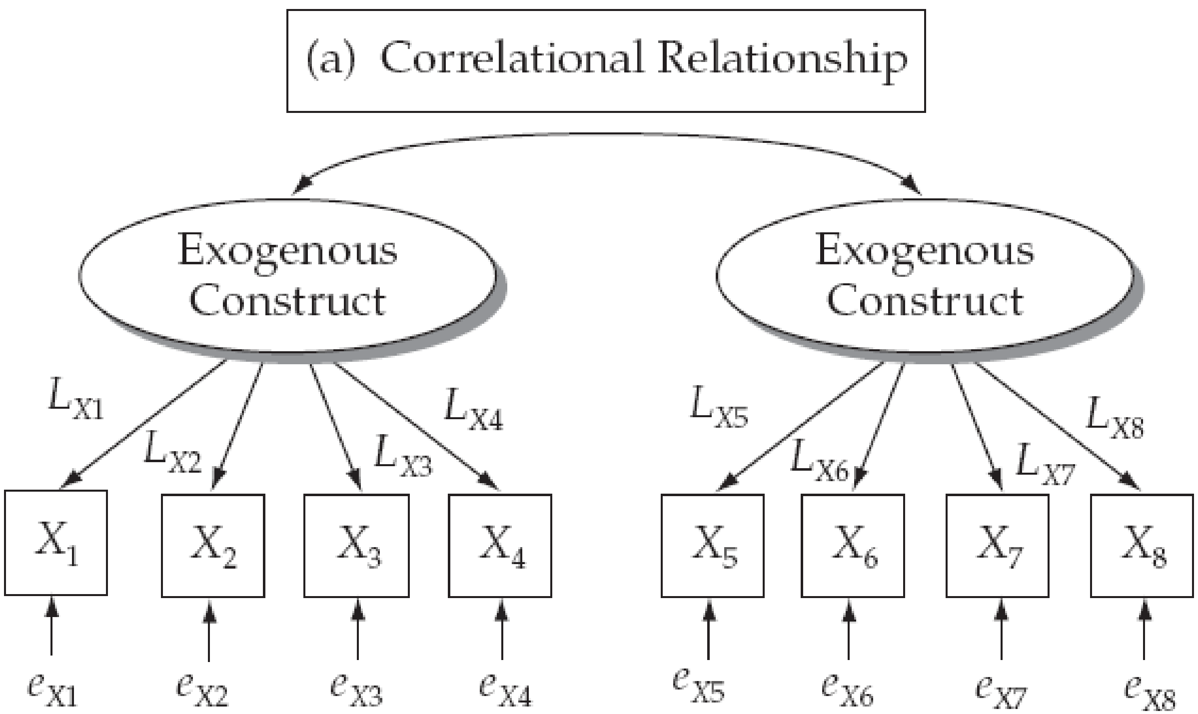
C. Dependence relationship between two constructs (structural relationship)



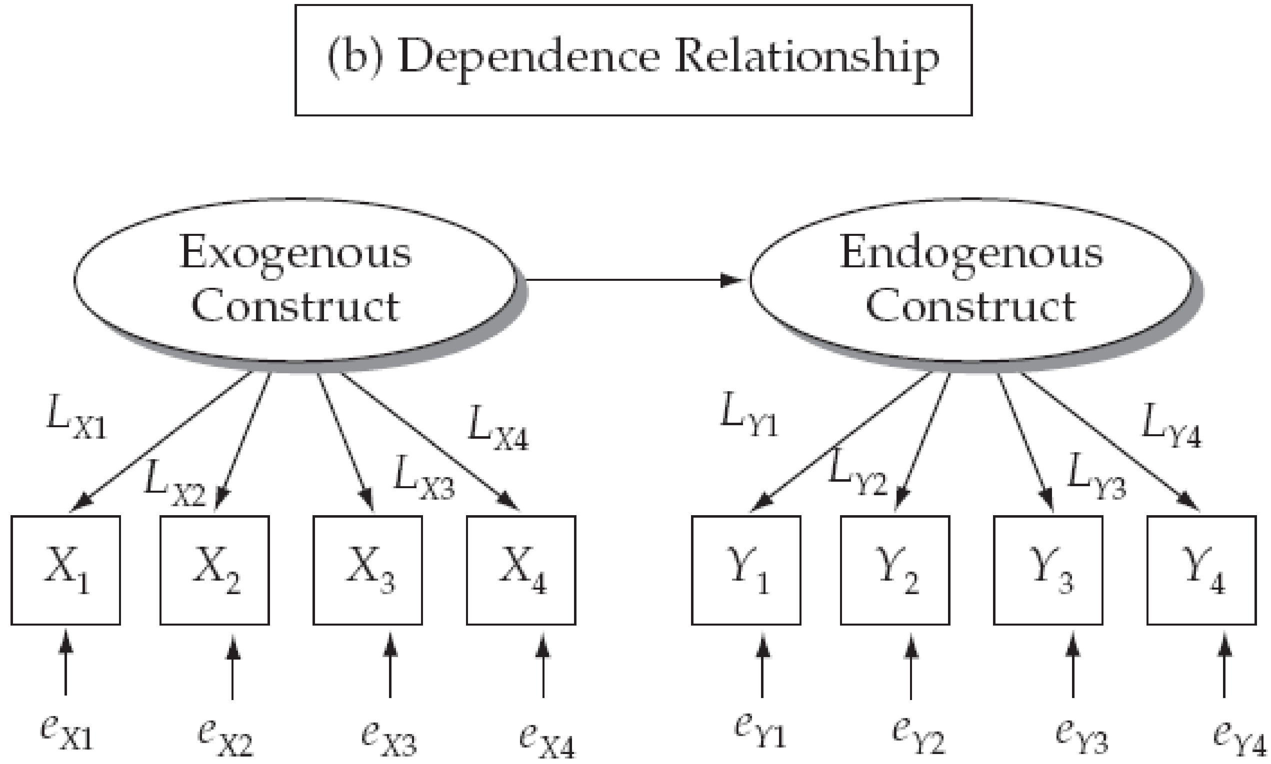
D. Correlational relationship between constructs



Visual Representation Measurement and Structural Model Relationships in a Simple SEM Model



Visual Representation Measurement and Structural Model Relationships in a simple SEM Model



The role of theory in SEM

- No model should be developed for use with SEM without some underlying theory.
- Theory is needed to develop both the
  - Measurement model specification and
  - Structural model specification
- SEM modeling strategies:
  - Confirmatory modeling strategy
  - Competing models strategy
  - Model development strategy

Causality

- Causal relationships
  - They are the strongest type of inference made in applying multivariate statistics.
  - They can be supported only when precise conditions for causality exist.

- Conditions for causality
  - Theoretical support exists for the relationship between the cause and effect.
  - Covariance between the cause and effect.
  - The cause must occur before the effect.
  - Nonspurious association must exist between the cause and effect.

Basics of SEM Estimation

SEM explains the *observed covariance* among a set of measured variables by comparing the observed covariance matrix with an *estimated covariance* matrix constructed based on the estimated relationships among variables.

Observed Covariance Matrix $\mathbf{S}$
Estimated Covariance Matrix $\mathbf{\Sigma}_k$

The closer these are, the better the fit. When they are equal, the fit is perfect.

Structural Equations Modeling Stages

Stage 1: Defining individual constructs

Stage 2: Developing the overall measurement model

Stage 3: Designing a study to produce empirical results

Stage 4: Assessing the measurement model validity

Stage 5: Specifying the structural model

Stage 6: Assessing structural model validity

Confirmatory Factor Analysis (CFA)

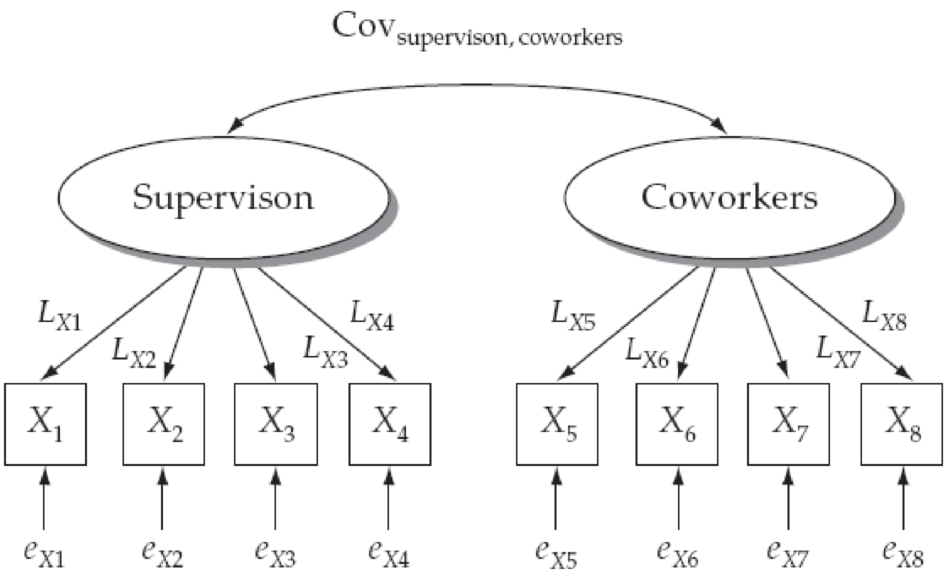
Structural Equation Modeling

Stage 1: Defining Individual Constructs

- Operationalizing the Constructs
- Scales from prior research vs. new scale development
- Pretesting

Stage 2: Developing the Overall Measurement Model

- Can the validity and unidimensionality of the constructs be supported?
- How many indicators for each construct?



Model-Identification

- Identification: Exists enough information to identify a solution for a set of structural equations?
  - Over-identified model: Model has more unique covariance and variance terms than parameters to estimate. It has positive degrees of freedom.
  - Just-identified model: SEM model containing just enough degrees of freedom to estimate all free parameters. Just identified models have perfect fit by definition, meaning a fit assessment is not meaningful.
  - Under-identified model: Model with more parameters to be estimated than there are item variance and covariance.

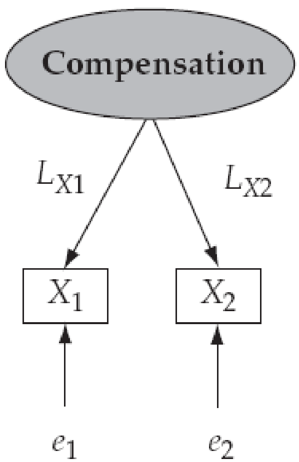
Model-Identification

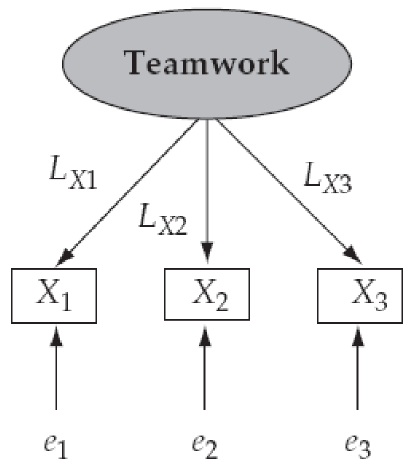
- Number of degrees of freedom: necessary condition

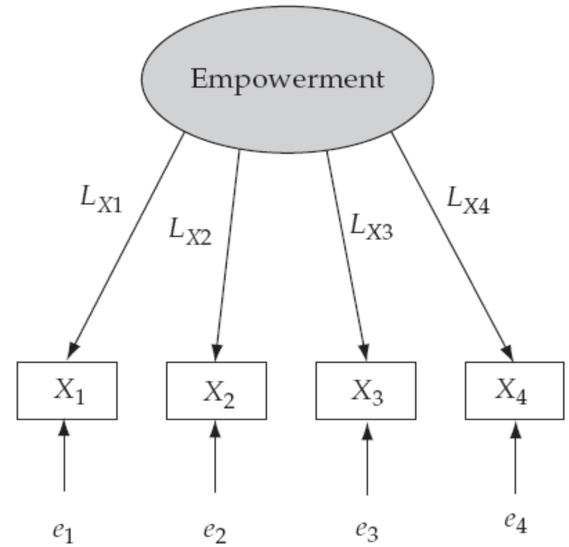
$$t \leq \frac{1}{2}(p+q)\cdot(p+q+1)$$

- t = number of parameters to be estimated
- p = number of y-variables
- q = number of x-variables

Model-Identification







S	X_1	X_2
X_1	**var(1)**	cov(1,2)
X_2	**cov(1,2)**	**var(2)**

S	X_1	X_2	X_3
X_1	**var(1)**	cov(1,2)	cov(1,3)
X_2	**cov(1,2)**	**var(2)**	cov(2,3)
X_3	**cov(1,3)**	**cov(2,3)**	**var(3)**

	X_1	X_2	X_3	X_4
X_1	2.01			
X_2	1.43	2.01		
X_3	1.31	1.56	2.24	
X_4	1.36	1.54	1.57	2.00

Model-Identification

Second requirement for identification:

- Every unobserved (latent) variable must be assigned a scale (metric)
- The ‘scale’ of a latent construct can be set by either:
  - Fixing one loading and setting its value to 1, or
  - Fixing the construct variance and setting its value to 1.

Stage 3: Designing a Study to Produce Empirical Results

Research Design

1. Type of data analyzes: covariances or correlations
2. Missing data
3. Sample size

Model Estimation

4. Model structure
5. Estimation techniques
6. Computer software used

Research Design

Type of data analyzes: covariances or correlations

- Today most SEM programs compute a model solution directly from raw data
- Covariance matrices provide the researcher with more flexibility due to the relatively greater information content and are the recommended form of input to SEM models.

Research Design

Missing data

- Use **list-wise** deletion with very large samples (i.e., respondent is eliminated)
- Use **pair-wise** deletion when the amount of missing data is less than 10 percent and the sample size is around 250 or more (i.e., all nonmissing data are used).
- Use one of the **imputation methods** for missing data as sample sizes become small or when missing data exceeds 10 percent (e.g., mean substitution)
- When the amount of missing data becomes very high (15 percent or more), SEM may not be appropriate.

Research Design

Sample size

- Depends on the model complexity and the communalities in each factor
- SEM models containing five or fewer constructs, each with more than three items (observed variables), and with high item communalities (.6 or higher), can be adequately estimated with samples as small as 100-150.
- When the number of factors is larger than six, some of which have fewer than three measured items as indicators, and multiple low communalities are present, sample size requirements may exceed 500.
- The sample size must be sufficient to allow the model to run, but more important, it must adequately represent the population.

Model Estimation

- Model structure: For each parameter, the researcher must decide if it is to be free or fixed
  - Fixed parameter: parameter that has a value specified by the researcher
  - Free parameter: parameter estimated by the structural equation program to represent the strength of a specified relationship
- Estimation techniques
- Computer software used

Stage 4: Assessing Measurement Model Validity

Finding specific evidence of construct validity

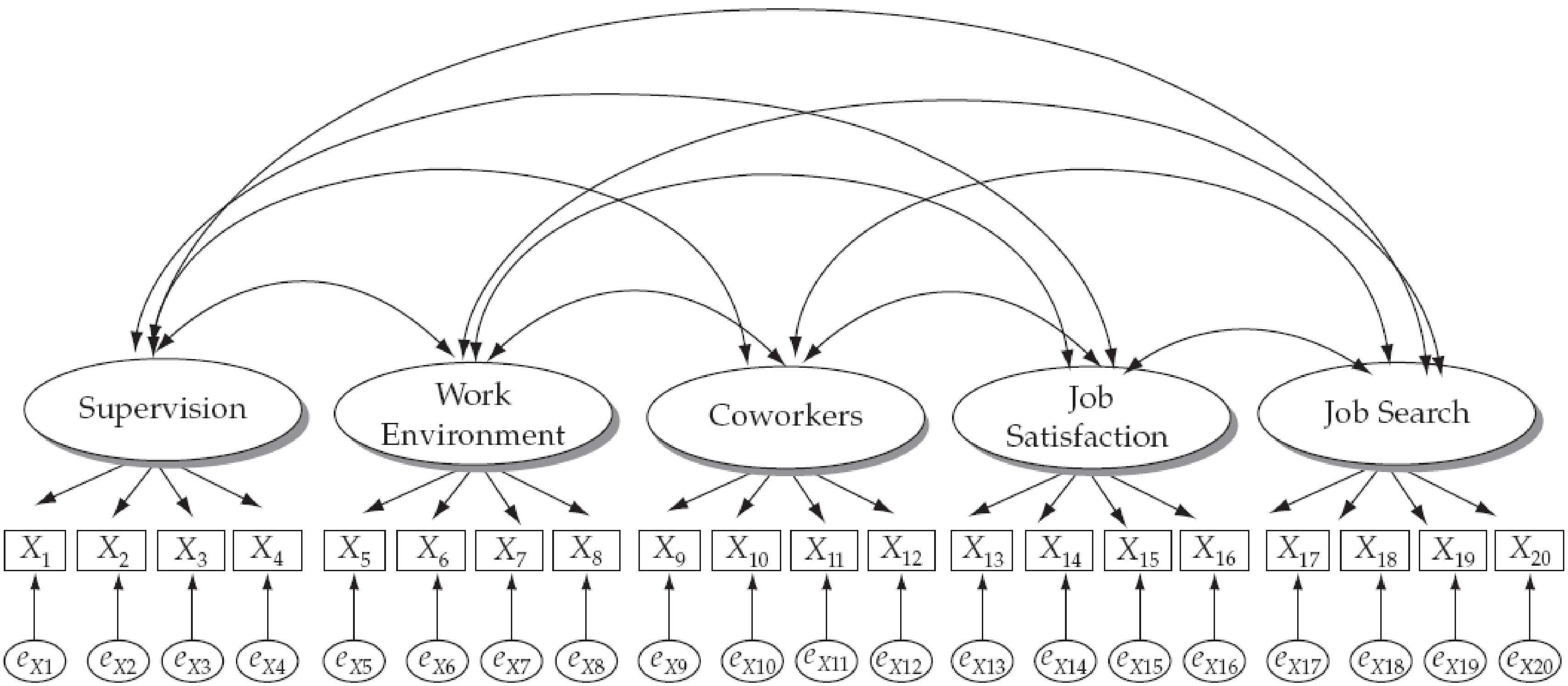
- Construct validity is the extent to which a set of measured items actually reflect the theoretical latent construct they are designed to measure.
- Confirmatory Factor Analysis (CFA), a special application of SEM, can be used to assess the construct validity of a proposed measurement theory
- Establishing acceptable levels of goodness-of-fit for the measurement model
- Goodness-of-fit (GOF): measure indicating how well a specified model reproduces the observed covariance matrix among the indicator variables

Confirmatory Factor Analysis vs. Exploratory Factor Analysis

- Confirmatory analysis: Use of a multivariate technique to test (confirm) a pre-specified relationship
- With CFA, the researcher must specify both
  - the number of factors that exist within a set of variables and
  - which factor each variable will load highly on

  before results can be computed. → theory driven
- With EFA,
  - all measured variables are related to *every* factor by a factor loading estimate
  - factors are derived from statistical results, not from theory → data driven

Hypothesized Measurement Model Specifications



Goodness-of-Fit-Indices

- What is goodness of fit (GOF)?

$$\chi^2 = (N-1)(S-\Sigma_k)$$

$$df = \frac{1}{2}[(p)(p+1)] - k$$

p = number of observed variables

k = number of estimated parameters

- Types of GOF:
  - Absolute Fit Measures.
  - Incremental Fit Measures.
  - Parsimonious Fit Measures.
- Evaluate construct validity of the measurement model.

Rules of Thumb: Assessing Predictive Accuracy

- As models become more complex, the likelihood of alternative models with equivalent fit increases.
- Multiple fit indices should be used to assess a model's goodness of fit. They should include:
  - The Chi-Square value and the associated df
  - One absolute fit index (like the GFI, RMSEA or SRMR)
  - One incremental fit index (like the CFI or TLI)
  - One goodness of fit index (GFI, CFI, TLI, . . .)
  - One badness of fit index (RMSEA, SRMR, . . .)

→ No single "magic" value for the fit indices that separates good from poor models.

Rules of Thumb: Assessing Predictive Accuracy

- It is not practical to apply a single set of cut-off rules that apply for all measurement models and for that matter to all SEM models of any type.

- The quality of fit depends heavily on model characteristics including sample size and model complexity.
  - Simple models with small samples should be held to very strict fit standards.
  - More complex models with larger samples should not be held to the same strict standards.

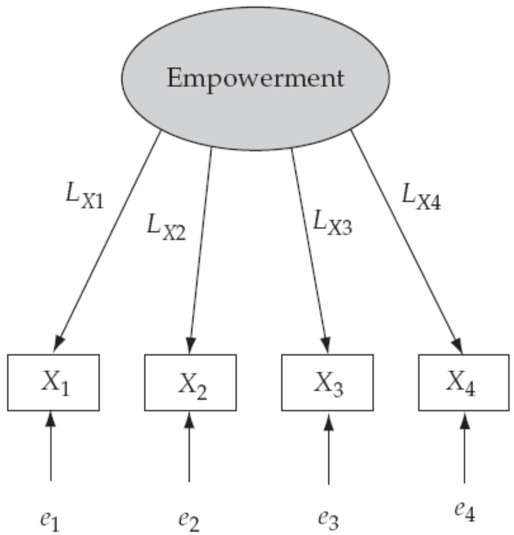
Guidelines for Establishing Acceptable and Unacceptable Fit

- Use multiple indices of differing types.
- Adjust the index cutoff values based on model characteristics.
- Use indices to compare models.
- The pursuit of better fit at the expense of testing a true model is not a good trade-off.

Rules of Thumb: Assessing Measurement Model Validity

- Loadings should be at least .5, ideally .7
- Items with low loadings become candidates for deletion
- Standardized loadings above +1.0 or below -1.0 are out of the feasible range and can be an important indicator of some problem with the data.

Example: Results of CFA



Eight paths to estimate

Symmetric Covariance Matrix

	X_1	X_2	X_3	X_4
X_1	2.01			
X_2	1.43	2.01		
X_3	1.31	1.56	2.24	
X_4	1.36	1.54	1.57	2.00

10 unique variance–covariance terms

Model Fit:
$\chi^2 = 14.9$
$df = 2$
$p = .001$
CFI = .99

Measured Items

X_1 = This organization allows me to do things I find personally satisfying.
X_2 = This organization provides an opportunity for me to excel in my job.
X_3 = I am encouraged to make suggestions about how this organization can be more effective.
X_4 = This organization encourages people to solve problems by themselves.

Loading Estimates	Error Variance Estimates
$L_{X1} = 0.78$	$e_1 = 0.39$
$L_{X2} = 0.89$	$e_2 = 0.21$
$L_{X3} = 0.83$	$e_3 = 0.31$
$L_{X4} = 0.87$	$e_4 = 0.24$

Assessing Measurement Model Validity

- Is the measurement model valid?

  No → refine measures and design a new study.

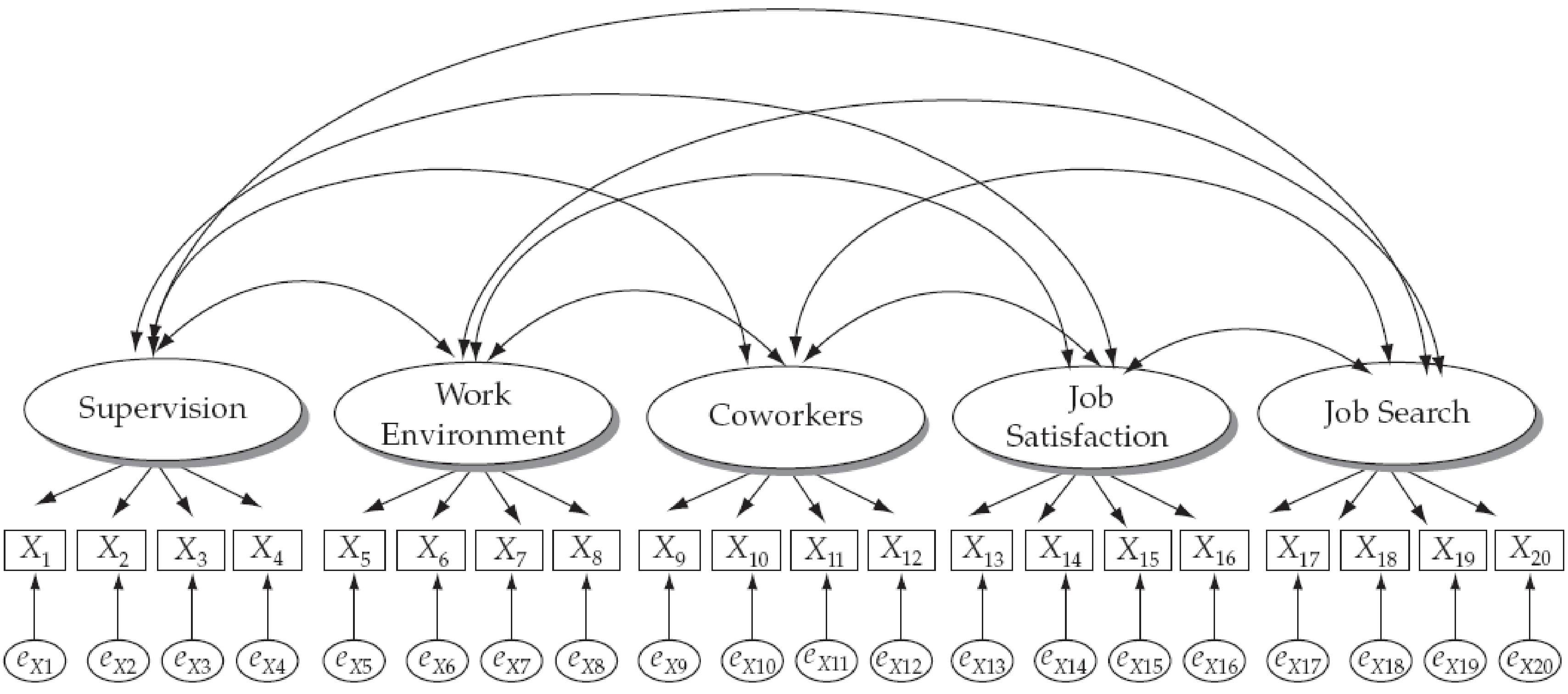
  Yes → proceed to test the structural model with stages 5 and 6.

Stage 5: Specifying the Structural Model

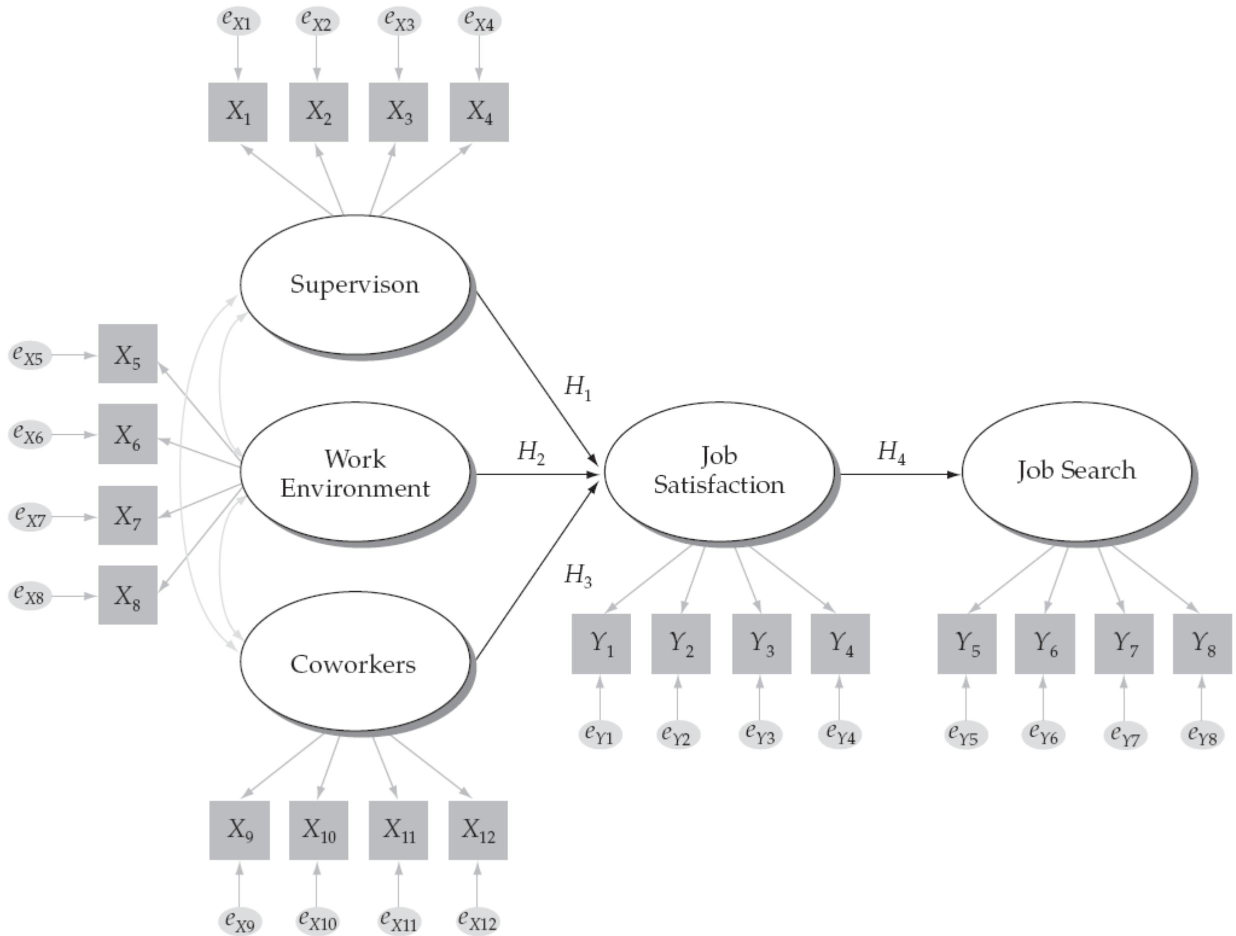
- A structural model should be tested after CFA has validated the measurement model.
  - assign relationships from one construct to another based on theory
  - specify the dependence relationships that exist among the constructs

- The structural relationships between constructs can be created by:
  - replacing the two-headed arrows from CFA with single headed arrows representing a cause and effect type relationship, or
  - removing the two-headed curved arrows connecting constructs that are not hypothesized to be directly related.

Hypothesized Measurement Model Specifications



Example: Specified Structural Model



Stage 6: Assessing the Structural Model Validity

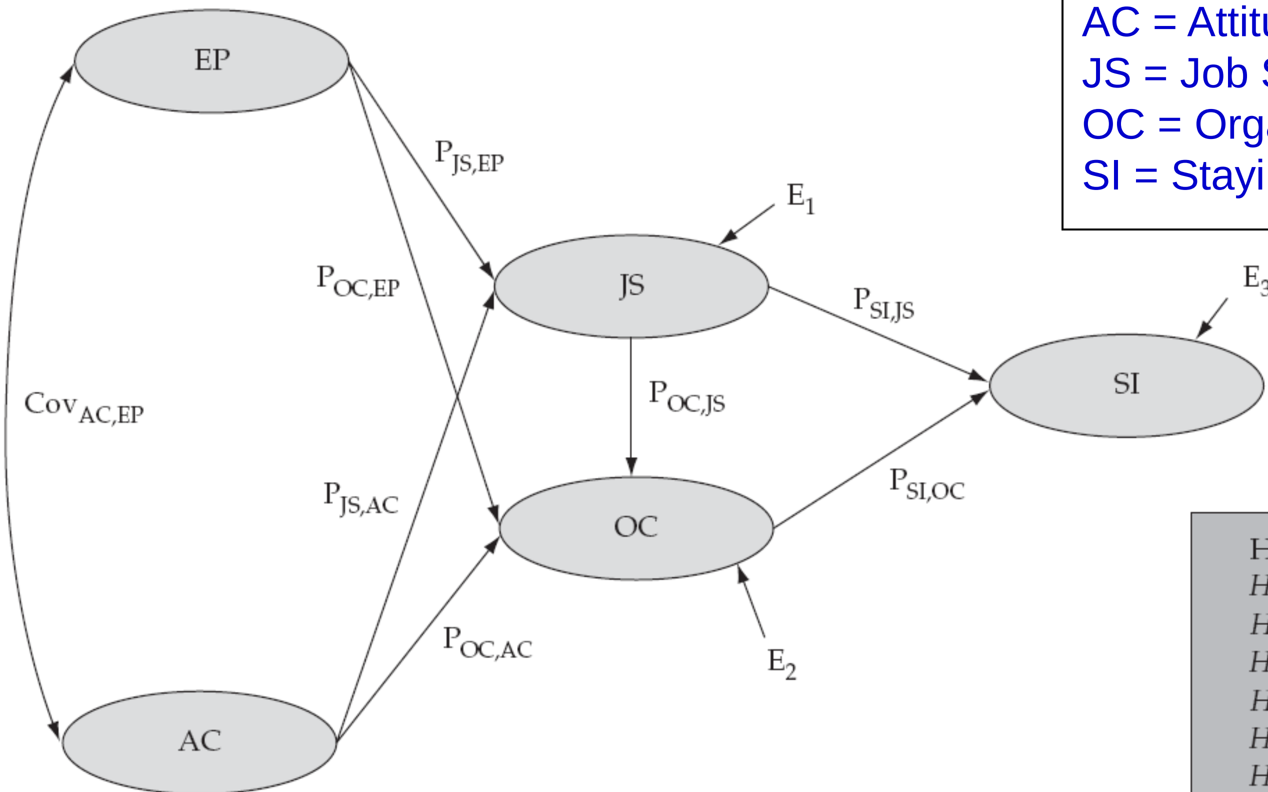
- Assess the goodness of fit (GOF) of the structural model.
- Evaluate the significance, direction, and size of the structural parameter estimates.

Is the structural model valid?

- No → refine model and test with new data.
- Yes → draw substantive conclusions and recommendations.

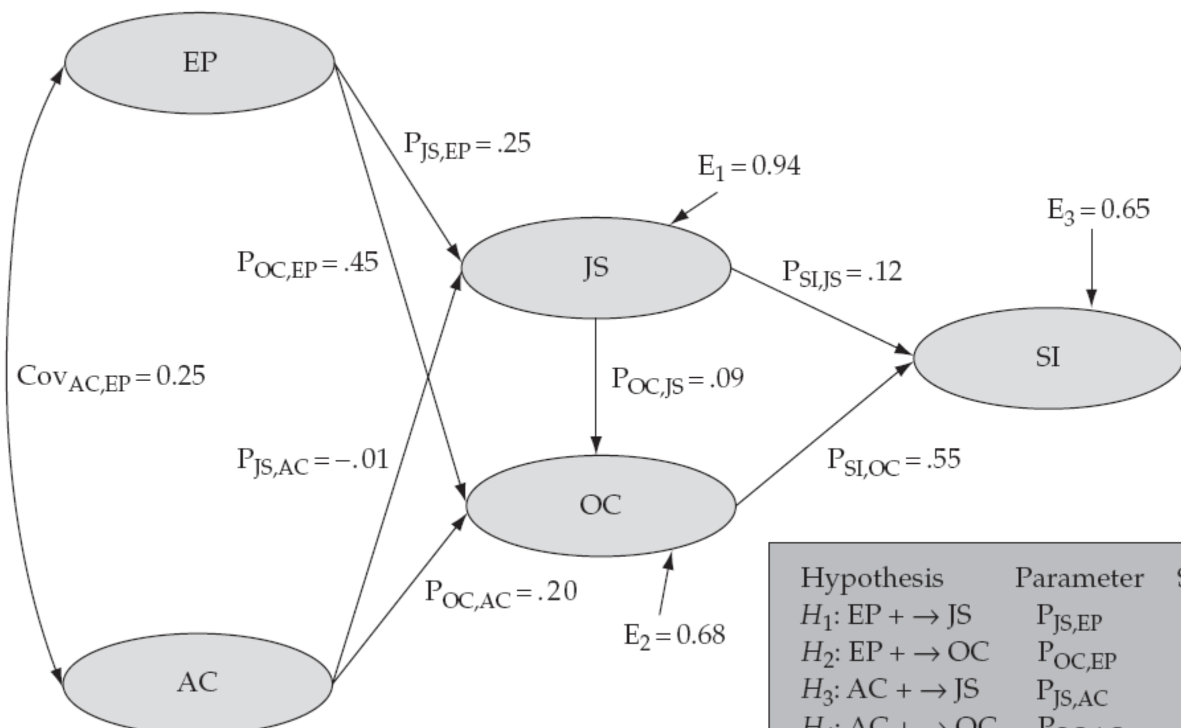
Example: The HBAT Employee Retention Model

EP = Environmental Perceptions
AC = Attitudes toward Co-workers
JS = Job Satisfaction
OC = Organizational Commitment
SI = Staying Intentions



Hypothesis	Parameter
H_1: EP + → JS	$P_{JS,EP}$
H_2: EP + → OC	$P_{OC,EP}$
H_3: AC + → JS	$P_{JS,AC}$
H_4: AC + → OC	$P_{OC,AC}$
H_5: JS + → OC	$P_{OC,JS}$
H_6: JS + → SI	$P_{SI,JS}$
H_7: OC + → SI	$P_{SI,OC}$

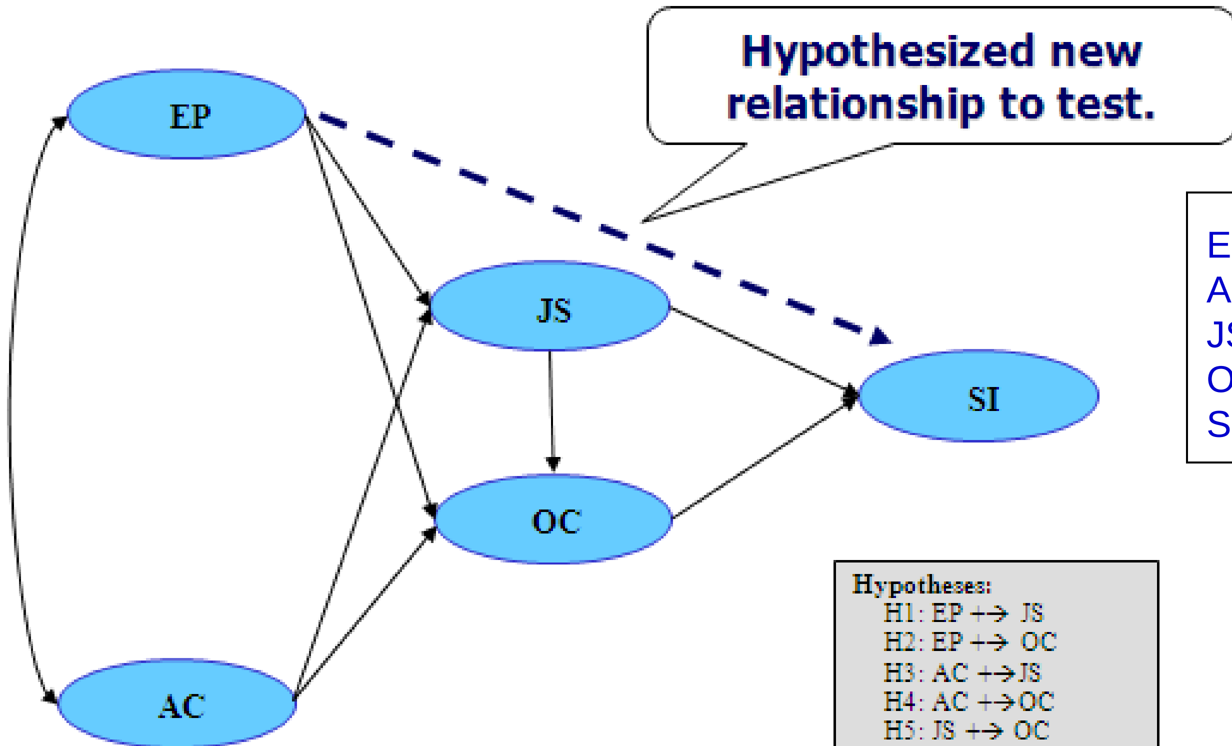
Example: Standardized Path Estimates for the model



EP = Environmental Perceptions
AC = Attitudes toward Co-workers
JS = Job Satisfaction
OC = Organizational Commitment
SI = Staying Intentions

Hypothesis	Parameter	Supported?
H_1: EP + → JS	$P_{JS,EP}$	Yes
H_2: EP + → OC	$P_{OC,EP}$	Yes
H_3: AC + → JS	$P_{JS,AC}$	No
H_4: AC + → OC	$P_{OC,AC}$	Yes
H_5: JS + → OC	$P_{OC,JS}$	No
H_6: JS + → SI	$P_{SI,JS}$	Yes
H_7: OC + → SI	$P_{SI,OC}$	Yes

Example: Possible "Competing" HBAT Model

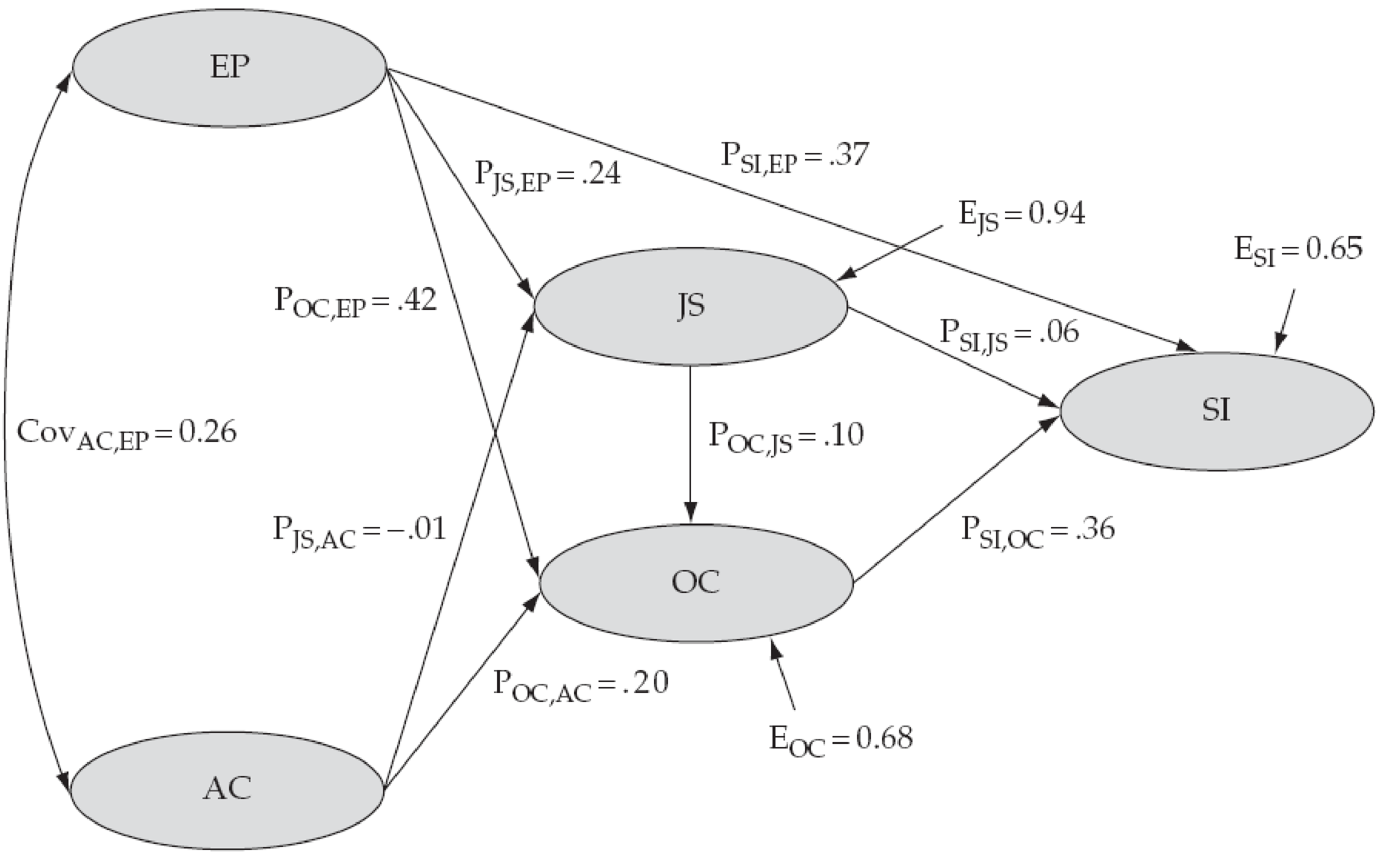


EP = Environmental Perceptions
AC = Attitudes toward Co-workers
JS = Job Satisfaction
OC = Organizational Commitment
SI = Staying Intentions

Hypotheses:

H1: EP +→ JS
H2: EP +→ OC
H3: AC +→ JS
H4: AC +→ OC
H5: JS +→ OC
H6: JS +→ SI
H7: OC +→ SI

Example: Standardized Path Estimates for the Revised Model



Example: GOF Measures for both models

GOF Index	Revised Employee Retention Model	Employee Retention Model
Absolute Measures		
χ^2 (chi-square)	242.23	283.43
Degrees of freedom	180	181
Probability	0.00	0.00
GFI	.95	.94
RMSEA	.029	.036
Confidence interval of RMSEA	.018–.038	.027–.045
RMR	.090	.110
SRMR	.040	.060
Normed chi-square	1.346	1.57
Incremental Fit Measures		
NFI	.97	.96
NNFI	.99	.98
CFI	.99	.99
RFI	.96	.96
Parsimony Measures		
AGFI	.93	.92
PNFI	.83	.83

Example: Comparison of the structural relationships in both models

HBAT Employee Retention Model		*Revised HBAT Employee Retention Model*	
Structural Relationship	**Standardized Parameter Estimate**	**Structural Relationship**	**Standardized Parameter Estimate**
H_1: EP → JS	0.25*	H_1: EP → JS	0.24*
H_2: EP → OC	0.45*	H_2: EP → OC	0.42*
H_3: AC → JS	–0.01	H_3: AC → JS	–0.01
H_4: AC → OC	0.20*	H_4: AC → OC	0.20*
H_5: JS → OC	0.09	H_5: JS → OC	0.10
H_6: JS → SI	0.12*	H_6: JS → SI	0.06
H_7: OC → SI	0.55*	H_7: OC → SI	0.36*
EP correlated AC	0.25*	EP correlated AC	0.26*
		EP → SI	0.37*

*Statistically significant at .05 level