

# Structural Equation Modeling (SEM) using R

Session 2: CFA and SEM

23-24 August 2021



# Topic overview

- 1: CFA-SEM overview
- 2: CFA-SEM with Lavaan
- 3: Defining constructs
- 4: Developing the overall measurement model
- 5: Assessing measurement model validity
- 6: Specifying the structural model
- 7: Assessing structural model validity

# CFA-SEM overview

---

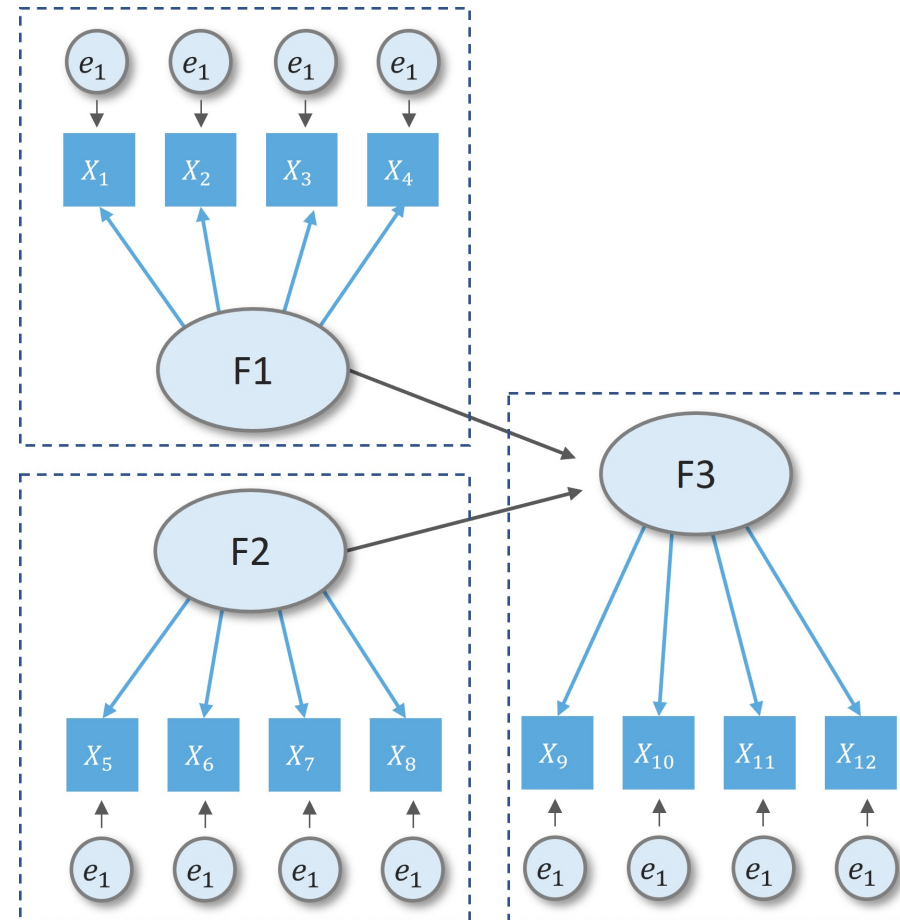
# What is SEM?

- Not a one statistical "technique"
- Integrates a number of different multivariate technique
  - Factor analysis
  - Regression
  - Simultaneous equation
- Distinction between:
  - measurement model
  - structural model

# What is SEM?

## Measurement model

- measurement part of a full SEM model
- confirmatory factor analysis



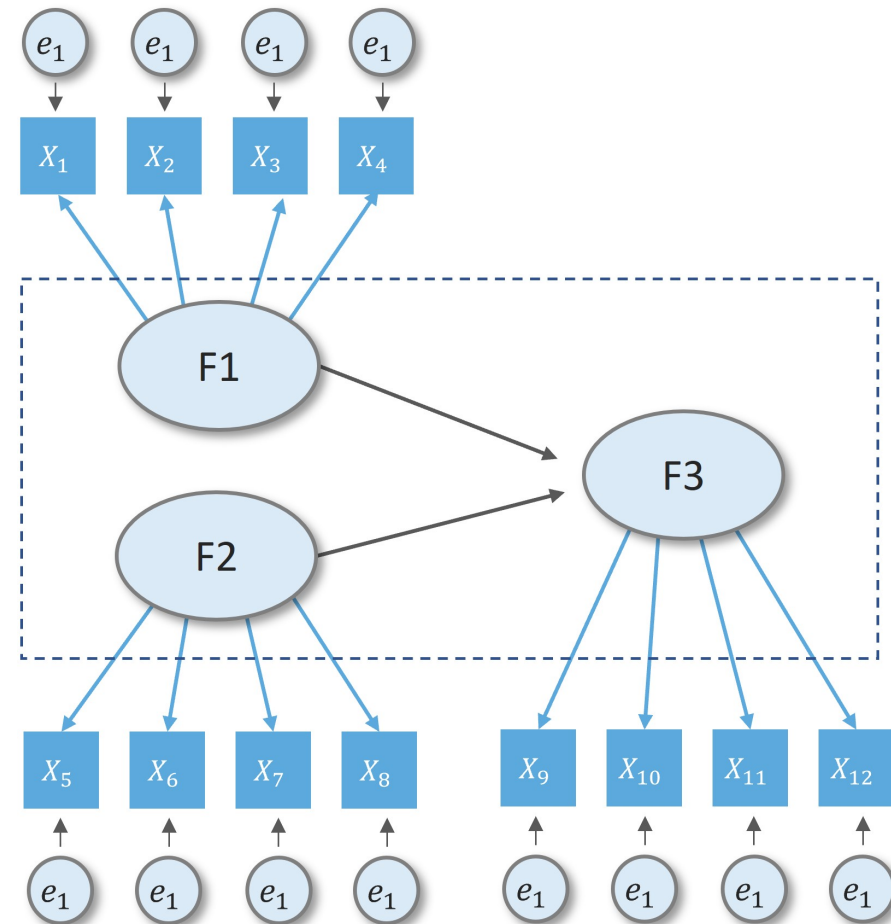
# What is SEM?

## Measurement model

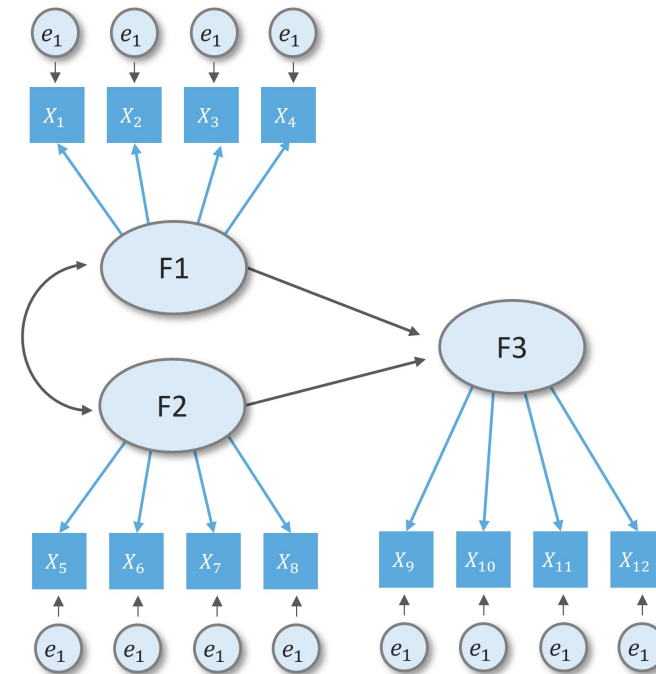
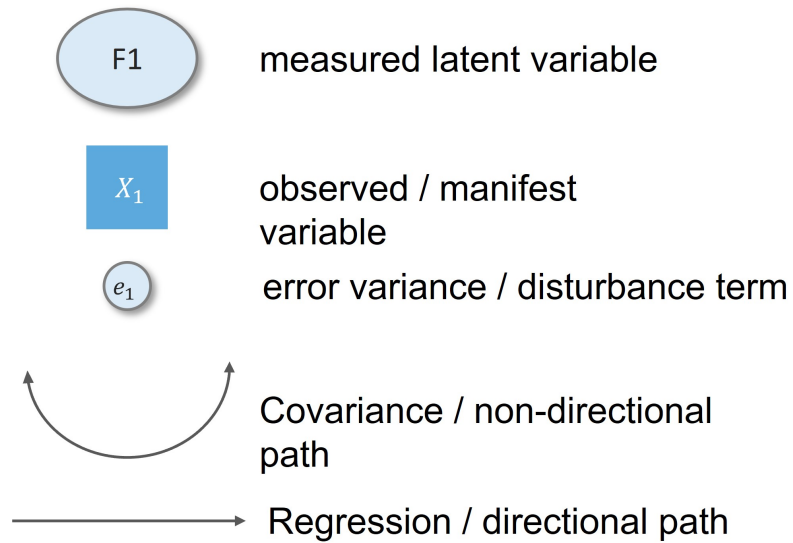
- measurement part of a full SEM model
- confirmatory factor analysis

## Structural model

- relationship between constructs
- full sem model is combination of measurement and structural component



# Basic SEM conventions



## 2. CFA-SEM with Lavaan R package

---



# What is Lavaan?

- *"developed to provide useRs, researchers, and teachers a free open-source, but commercial quality", Yves Rosseel (2012)*
- Check-out their [lavaan tutorial](#)

```
install.packages("lavaan")  
library(lavaan)  
example(cfa)
```

```
cfa> ## The famous Holzinger and Swineford (1939) example  
cfa> HS.model <- ' visual  =~ x1 + x2 + x3  
cfa+                textual =~ x4 + x5 + x6  
cfa+                speed   =~ x7 + x8 + x9 '
```

```
cfa> fit <- cfa(HS.model, data = HolzingerSwineford1939)
```

```
cfa> summary(fit, fit.measures = TRUE)  
lavaan 0.6-9 ended normally after 35 iterations
```

Estimator	ML
Optimization method	NLMINB
Number of model parameters	21

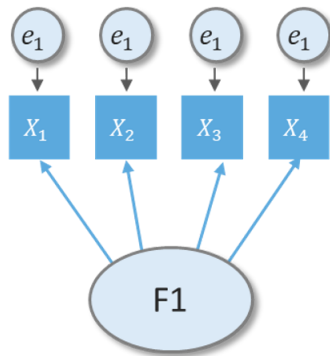
# Major operators of lavaan syntax

Command	Operator	Illustration	Significance
Estimate covariance	~~	X ~~ Y	X is correlated with Y
Estimate regression	~	Y ~ X	Y is regressed on X
Define a reflective latent variable	= ~	F =~ item_1 + item_2 + item_3	The F factor is measured by indicators item 1, item 2, and item 3 over which it has effects
Label a parameter	*	F =~ b1*item_1 + b2*item2 + b3*item3	Item 1-3 is named “b1”, “b2”, and “b3”, respectively.
Create a new parameter	:=	B1b2 := b1*b2	Define a parameter that is not in the model. For example: b1b2 = indirect effect of b1 and b2
Insert a comment in the syntax	#	#indirect effects B1b2 := b1*b2	Explain to the reader the meaning of a command.

# Major operators of lavaan syntax

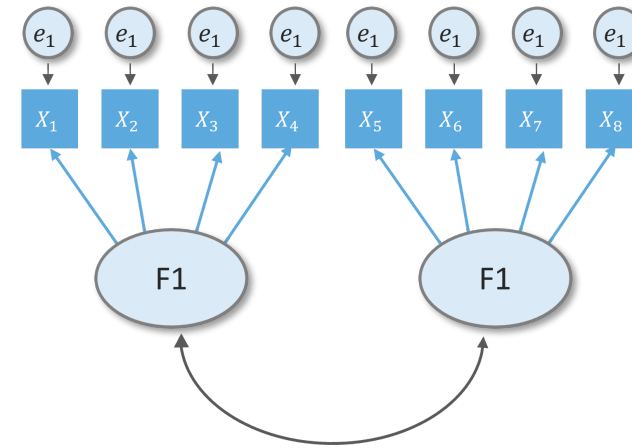
## Defining a reflective latent variable

```
model <- "F1 =~ x1 + x2 + x3 + x4"
```



## Estimate factor covariance

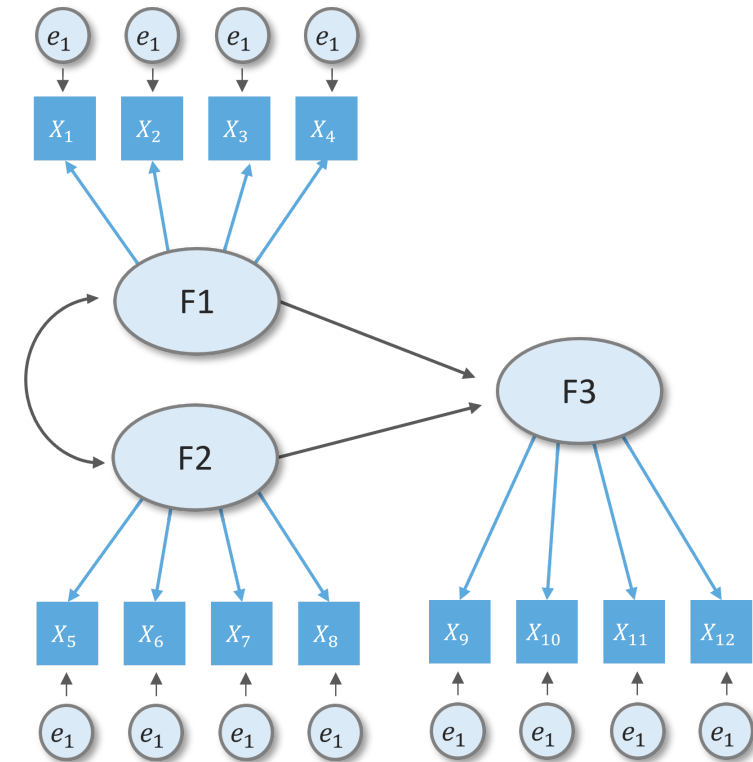
```
model <- "F1 =~ x1 + x2 + x3 + x4  
          F2 =~ x5 + x6 + x6 + x8  
          F1 ~~ F2"
```



# Major operators of lavaan syntax

## Estimate regression

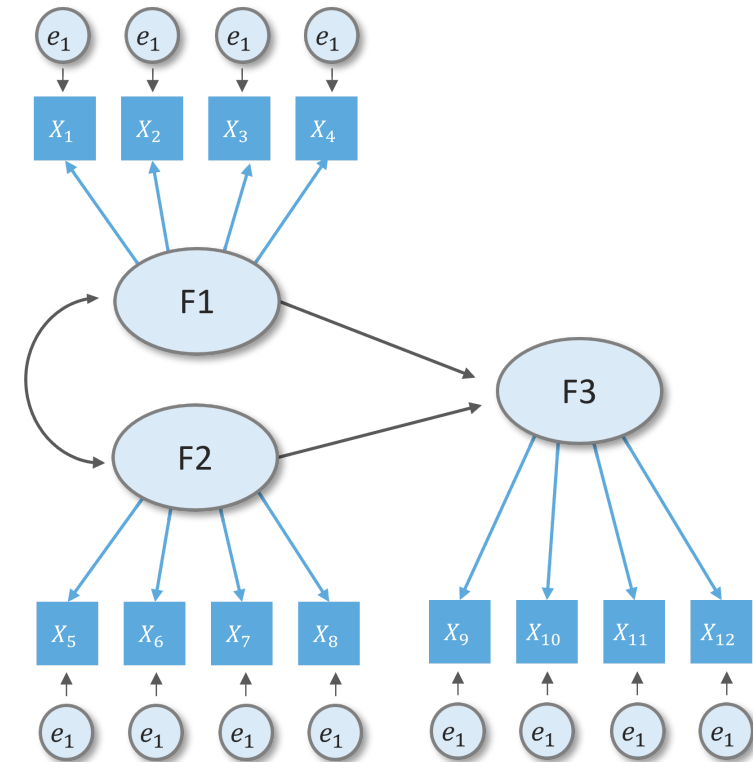
```
model <- "F1 =~ x1 + x2 + x3 + x4  
         F2 =~ x5 + x6 + x7 + x8  
         F3 =~ x9 + x10 + x11 + x12  
         F1 ~~ F2  
         F3 ~ F1 + F2"
```



# Major operators of lavaan syntax

## Insert a comment in the syntax

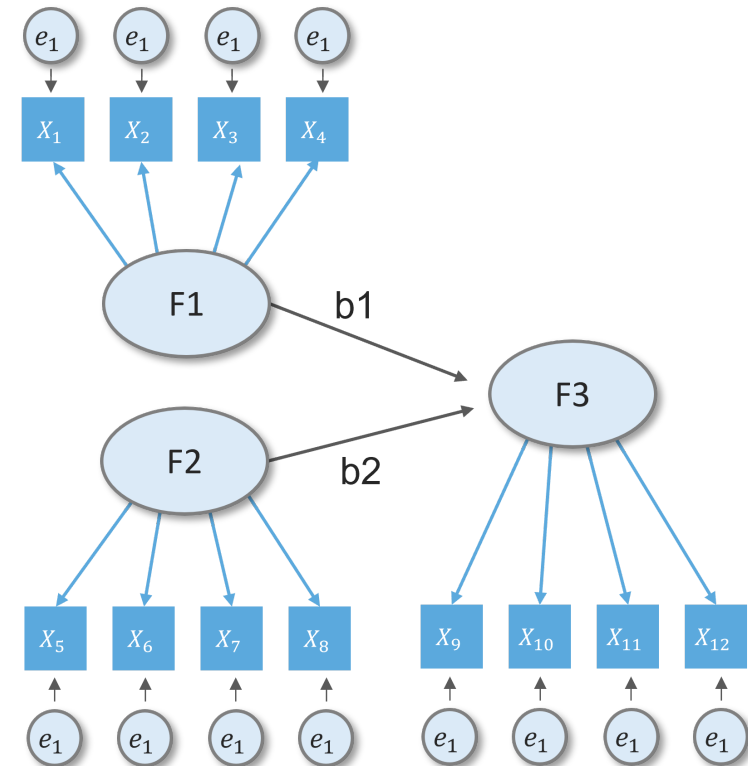
```
model <- "F1 =~ x1 + x2 + x3 + x4  
         F2 =~ x5 + x6 + x7 + x8  
         F3 =~ x9 + x10 + x11 + x12  
  
         # covariance  
         F1 ~~ F2  
  
         # F3 is regressed on F1 and F2  
         F3 ~ F1 + F2"
```



# Major operators of lavaan syntax

## Label a parameter

```
model <- "F1 =~ x1 + x2 + x3 + x4  
         F2 =~ x5 + x6 + x7 + x8  
         F3 =~ x9 + x10 + x11 + x12  
  
         # covariance  
         F1 ~~ F2  
  
         # F3 is regressed on F1 and F2  
         F3 ~ b1*F1 + b2*F2"
```



# Major operators of lavaan syntax

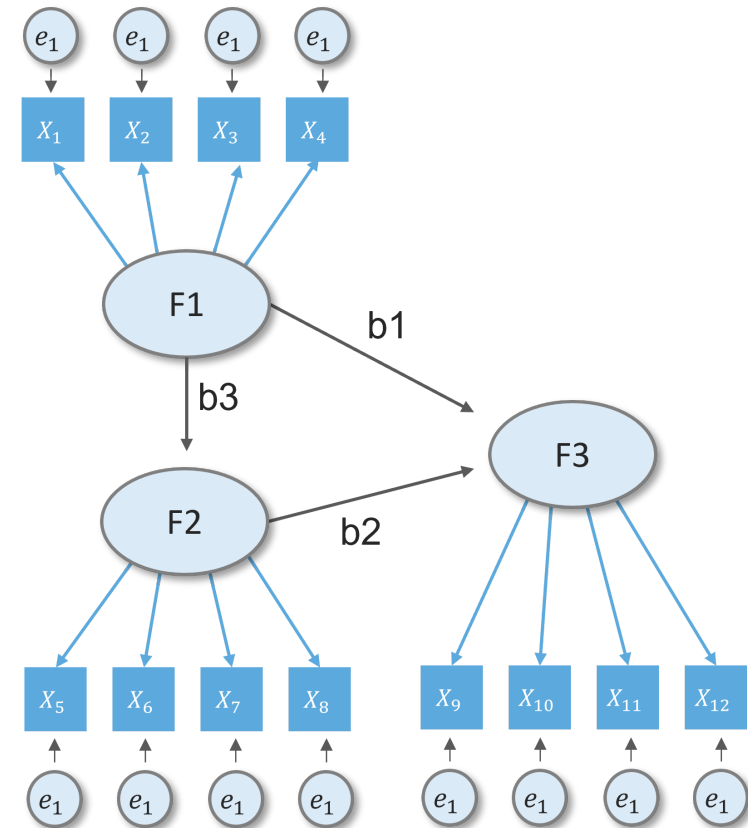
## Create a new parameter

```
model <- "F1 =~ x1 + x2 + x3 + x4
         F2 =~ x5 + x6 + x7 + x8
         F3 =~ x9 + x10 + x11 + x12

         # regression
         F3 ~ b1*F1 + b2*F2
         F2 ~ b3*F1

         # F1 indirect effect
         ie := b3*b2

         # F1 total effect
         te := b3*b2 + b1"
```



# Main steps in SEM

---



# Main steps in SEM

1. Defining constructs
2. Developing the overall measurement model
3. Assessing measurement model validity
4. Specifying the structural model
5. Assessing structural model validity

# 1. Defining Constructs

---

# Dataset

- HBAT company
- HBAT is interested in understanding what affects employee's attitudes and behaviors that contributes to employee's retention.

JS1	OC1	OC2	EP1	OC3	OC4	EP2	EP3	AC1	EP4
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
5	3	5	10	10	10	10	5	1	2
3	0	5	10	3	7	10	10	2	7
4	6	10	10	10	10	10	10	1	7
4	7	7	10	10	7	10	9	2	7
5	2	10	10	9	9	9	10	1	6
6	5	8	8	7	7	10	7	1	7
2	6	10	9	10	9	9	9	2	6
2	4	9	10	9	7	10	10	1	7
4	9	10	8	10	10	6	8	3	3
5	5	9	10	9	10	10	8	2	7
1-10 of 400 rows   1-1... Previous 1 2 3 4 5 6 ... 40 Next									

# Defining individual constructs

- Based on literature and preliminary interviews, a study was designed focusing on five key constructs.
  - *Job satisfaction (JS)* : reactions resulting from an appraisal of one's job situation.
  - *Organizational commitment (OC)*: extent to which an employees indentifies and feels part of HBAT.
  - *Staying intention (SI)*: extent to which an employee intends to continue working for HBAT.
  - *Environmental perceptions (EP)*: beliefs an employee has about day-to-day, physical working conditions.
  - *Attitudes towards coworkers (AC)*: attitudes an employee has toward the coworkers he/she interacts with on a regular basis.

# Defining individual constructs

Item	Scale Type	Description	Construct
JS <sub>1</sub>	0–10 Likert Disagree–Agree	All things considered, I feel very satisfied when I think about my job.	JS
OC <sub>1</sub>	0–10 Likert Disagree–Agree	My work at HBAT gives me a sense of accomplishment.	OC
OC <sub>2</sub>	0–10 Likert Disagree–Agree	I am willing to put in a great deal of effort beyond that normally expected to help HBAT be successful.	OC
EP <sub>1</sub>	0–10 Likert Disagree–Agree	I am comfortable with my physical work environment at HBAT.	EP
OC <sub>3</sub>	0–10 Likert Disagree–Agree	I have a sense of loyalty to HBAT.	OC
OC <sub>4</sub>	0–10 Likert Disagree–Agree	I am proud to tell others that I work for HBAT.	OC
EP <sub>2</sub>	0–10 Likert Disagree–Agree	The place I work in is designed to help me do my job better.	EP
EP <sub>3</sub>	0–10 Likert Disagree–Agree	There are few obstacles to make me less productive in my workplace.	EP
AC <sub>1</sub>	5-point Likert	How happy are you with the work of your coworkers? ___ Not happy ___ Somewhat happy ___ Happy ___ Very happy ___ Extremely happy	AC
EP <sub>4</sub>	7-point Semantic Differential	What term best describes your work environment at HBAT? Too hectic _____ Very soothing	EP
JS <sub>2</sub>	7-point Semantic Differential	When you think of your job, how satisfied do you feel? Not at all satisfied _____ Very much satisfied	JS
JS <sub>3</sub>	7-point Semantic Differential	How satisfied are you with your current job at HBAT? Very unsatisfied _____ Very satisfied	JS
AC <sub>2</sub>	7-point Semantic Differential	How do you feel about your coworkers? Very unfavorable _____ Very favorable	AC
SI <sub>1</sub>	5-point Likert Disagree–Agree	I am not actively searching for another job. Strongly disagree _____ Strongly agree	SI
JS <sub>4</sub>	5-point Likert	How satisfied are you with HBAT as an employer? ___ Not at all ___ Little ___ Average ___ A lot ___ Very much	JS
SI <sub>2</sub>	5-point Likert Disagree–Agree	I seldom look at the job listings on monster.com. Strongly disagree _____ Strongly agree	SI
JS <sub>5</sub>	Percent Satisfaction	Indicate your satisfaction with your current job at HBAT by placing a percentage in the blank, with 0% = Not satisfied at all, and 100% = Highly satisfied. _____	JS
AC <sub>3</sub>	5-point Likert	How often do you do things with your coworkers on your days off? Never ___ Rarely ___ Occasionally ___ Often ___ Very often	AC

AC <sub>3</sub>	5-point Likert	How often do you do things with your coworkers on your days off? ___ AC Never ___ Rarely ___ Occasionally ___ Often ___ Very often	AC
SI <sub>3</sub>	5-point Likert Disagree–Agree	I have no interest in searching for a job in the next year. Strongly disagree _____ Strongly agree	SI
AC <sub>4</sub>	6-point Semantic Differential	Generally, how similar are your coworkers to you? Very different _____ Very similar	AC
SI <sub>4</sub>	5-point Likert	How likely is it that you will be working at HBAT one year from today? ___ Very unlikely ___ Unlikely ___ Somewhat likely ___ Likely ___ Very likely	SI

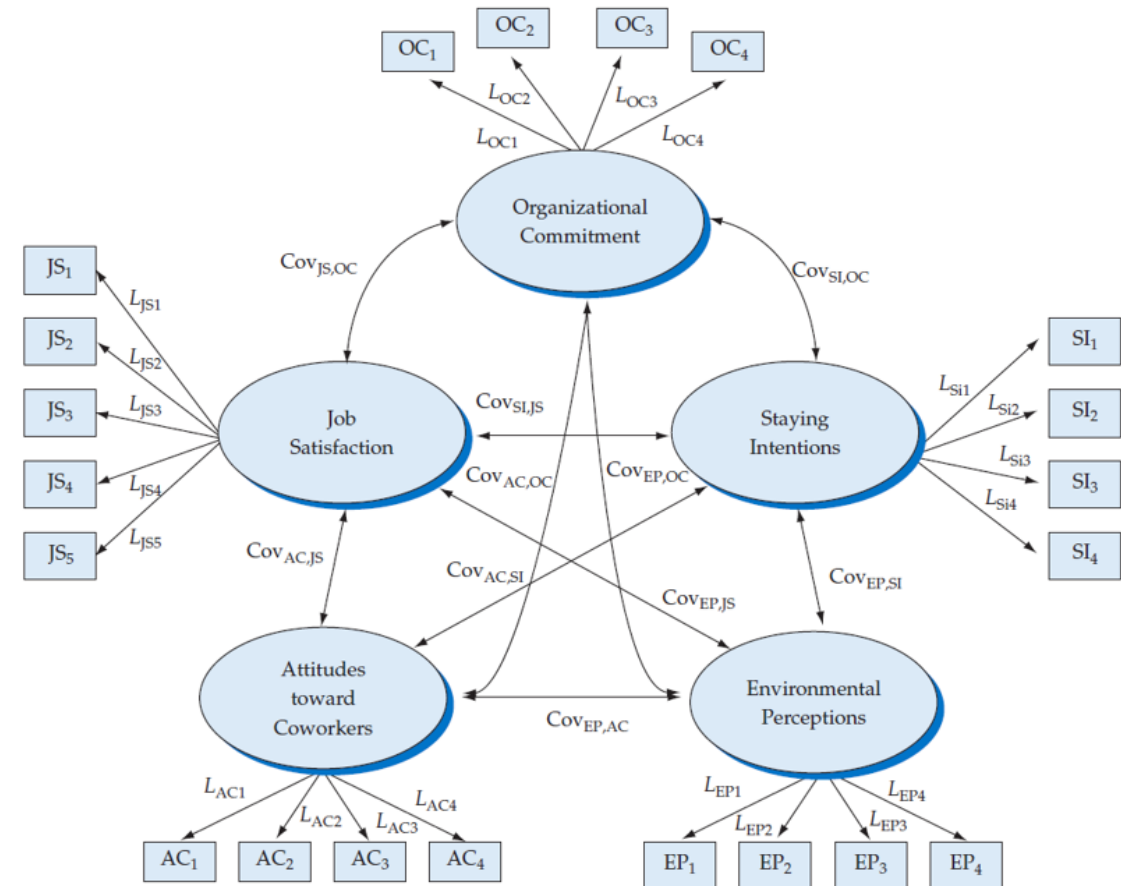
Source: JF Hair et al. (2019) : Multivariate data analysis

## Step 2. Developing overall measurement model

---

# Developing overall measurement model

- Measurement theory model (CFA) for HBAT employees
- Direction of the relationship between factors is not yet defined.
- Focus on confirming the specified model with empirical model (using empirical data), hence confirmatory.



**Let's practice!**



# Step 3. Assessing measurement model validity

---

# Basic principles

- Compare covariance matrix of the research data  $S$  and reproduced covariance  $\Sigma$
- Hypothesis:
  - Null:  $S = \Sigma$
  - Alternative:  $S \neq \Sigma$
- Idea is to arrived with a parameter that minimizes the difference of  $S$  and  $\Sigma$

```
cfa_fit <- cfa(cfa_model, data = hbat_data)
cfa_fit %>% summary()
```

```
lavaan 0.6-9 ended normally after 54 iterations
```

Estimator	ML
Optimization method	NLMINB
Number of model parameters	52
Number of observations	400

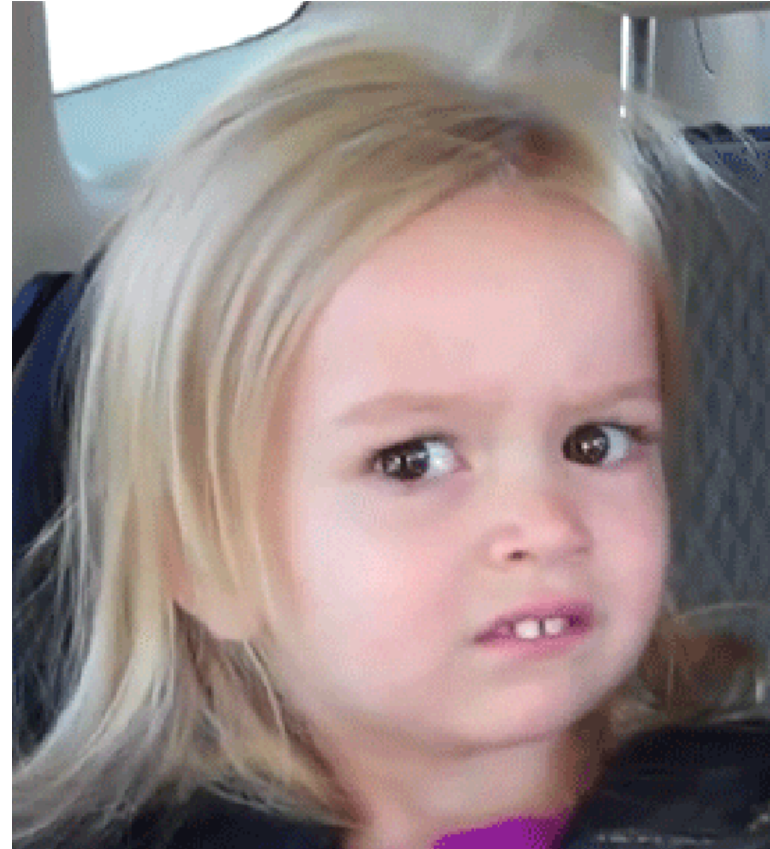
```
Model Test User Model:
```

Test statistic	240.738
Degrees of freedom	179
P-value (Chi-square)	0.001

```
Parameter Estimates:
```

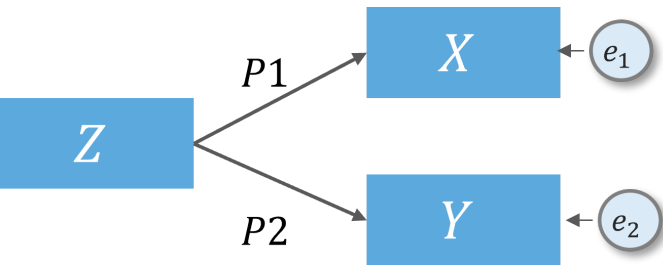
# Basic principles

- Compare covariance matrix of the research data  $S$  and reproduced covariance  $\Sigma$
- Hypothesis:
  - Null:  $S = \Sigma$
  - Alternative:  $S \neq \Sigma$
- Idea is to arrived with a parameter that minimizes the difference of  $S$  and  $\Sigma$



# Basic principles

- To understand the SEM process, consider the Table on the right.
- e.g., iterative procedure using least square method.



	Parameters		$r_{XZ} = 0.79$	$r_{YZ} = 0.59$	$r_{XY} = 0.49$	$\sum d^2$
Iteration cycles	P1	P2	Reproduced correlations			Least squares
1	0.50	0.50	0.50	0.50	0.250	0.149
1a	0.49	0.49	0.49	0.49	0.240	0.162
1b	0.49	0.50	0.49	0.50	0.245	0.158
1c	0.50	0.49	0.50	0.49	0.245	0.158
2	0.55	0.55	0.55	0.55	0.300	0.094
2a	0.60	0.60	0.60	0.60	0.360	0.024
3	0.65	0.61	0.65	0.61	0.400	0.027
3a	0.65	0.62	0.65	0.62	0.403	0.028
4	0.67	0.61	0.67	0.61	0.408	0.021
4a	0.70	0.61	0.70	0.61	0.427	0.012
4b	0.75	0.61	0.75	0.61	0.457	0.003
4c	0.80	0.61	0.80	0.61	0.480	0.0006
5	0.81	0.61	0.81	0.61	0.494	0.0008

# Summary output

- Overall results
- Loadings
- Variances

```
cfa_fit <- cfa(cfa_model, data = hbat_data)
summary(cfa_fit)
```

```
lavaan 0.6-9 ended normally after 54 iterations
```

Estimator	ML
Optimization method	NLMINB
Number of model parameters	52
Number of observations	400

```
Model Test User Model:
```

Test statistic	240.738
Degrees of freedom	179
P-value (Chi-square)	0.001

```
Parameter Estimates:
```

# Summary output

## Overall results

- Degrees of freedom (df)
  - $df = \frac{1}{2}p(p + 1) - k$
  - $p$  = total observed variables
  - $k$  = total estimated parameters
- Identification
  - Include at least three manifest variables
  - Create models with  $df > 0$

```
cfa_fit <- cfa(cfa_model, data = hbat_data)
summary(cfa_fit)
```

```
lavaan 0.6-9 ended normally after 54 iterations
```

Estimator	ML
Optimization method	NLMINB
Number of model parameters	52
Number of observations	400

```
Model Test User Model:
```

Test statistic	240.738
Degrees of freedom	179
P-value (Chi-square)	0.001

```
Parameter Estimates:
```

# Summary output

## Loadings

- Measures the strength of the relationship between items and factor.

```
cfa_fit <- cfa(cfa_model, data = hbat_data)
summary(cfa_fit, standardized = TRUE)
```

```
lavaan 0.6-9 ended normally after 54 iterations
```

Estimator	ML
Optimization method	NLMINB
Number of model parameters	52
Number of observations	400

```
Model Test User Model:
```

Test statistic	240.738
Degrees of freedom	179
P-value (Chi-square)	0.001

```
Parameter Estimates:
```

# Summary output

## Variances

Refer to unique variance that the factor unable to account for. Similar to error term in OLS, hence it is also term as error variance.

```
cfa_fit <- cfa(cfa_model, data = hbat_data)
summary(cfa_fit, standardized = TRUE)
```

```
lavaan 0.6-9 ended normally after 54 iterations
```

Estimator	ML
Optimization method	NLMINB
Number of model parameters	52
Number of observations	400

```
Model Test User Model:
```

Test statistic	240.738
Degrees of freedom	179
P-value (Chi-square)	0.001

```
Parameter Estimates:
```



# Fit indices

## Goodness of fit indices

- Goodness-of-fit index (GFI)
- Adjusted goodness-fit-index (AGFI)
- Comparative fit index (CFI)
- Normed fit index (NFI)
- Non-normed fit index (NNF)

## Badness of fit indices

- Standard root mean square of the residuals (SRMR)
- Root mean square error of approximation (RMSEA)

Table 3. Goodness of fit of the measurement model.

Fit indices	Recommended value	Sources	Research model
$\chi^2$	—	—	369.4
df	—	—	120
$\chi^2/\text{df}$	< 5	Bollen (1989)	3.08
GFI	> 0.9	Scott (1995)	0.91
AGFI	> 0.8	Scott (1995)	0.87
SRMR	< 0.1	Hu and Bentler (1999)	0.034
CFI	> 0.9	Bagozzi and Yi (1988)	0.96
RMSEA	< 0.08	MacCallum <i>et al.</i> (1996)	0.071
NFI	> 0.9	Bentler and Bonett (1980)	0.95
NNFI	> 0.9	Bentler and Bonett (1980)	0.95

*Sample GOF results from W. Shiau & M. Luo (2013). Continuance intention of blog users: The impact of perceived enjoyment, habit, user involvement and blogging time.*

# Fit indices

## Goodness of fit indices

- Goodness-of-fit index (GFI)
- Adjusted goodness-fit-index (AGFI)
- Comparative fit index (CFI)
- Normed fit index (NFI)
- Non-normed fit index (NNF)

```
fitMeasures(cfa_fit)
```

npars	fmin	chisq
52.000	0.301	240.738
pvalue	baseline.chisq	baseline.df
0.001	4452.567	210.000
cfi	tli	nnfi
0.985	0.983	0.983
nfi	pnfi	ifit
0.946	0.806	0.980
logl	unrestricted.logl	aic
-13916.782	-13796.413	27937.564
ntotal	bic2	rmsea
400.000	27980.120	0.029
rmsea.ci.upper	rmsea.pvalue	rmr
0.039	1.000	0.414
srmr	srmr_bentler	srmr_bentler_nomean
0.036	0.036	0.030
crmr_nomean	srmr_mplus	srmr_mplus_nomean
0.037	0.036	0.030
cn_01	gfi	agfi
376.401	0.947	0.933

# Fit indices

## Goodness of fit indices

- Goodness-of-fit index (GFI)
- Adjusted goodness-fit-index (AGFI)
- Comparative fit index (CFI)
- Normed fit index (NFI)
- Non-normed fit index (NNF)

```
fitMeasures(cfa_fit, fit.measures = c("gfi", "agfi", "cfi"
```

```
  gfi  agfi  cfi  nfi  nnfi  
0.947 0.932 0.985 0.946 0.983
```

# Fit indices

## Badness of fit indices

- Standard root mean square residual (SRMR)
- Root mean square error of approximation (RMSEA)

```
fitMeasures(cfa_fit, fit.measures = c("srmr", "rmsea"))
```

```
srmr rmsea  
0.036 0.029
```

# Reliability and validity test

## Reliability test

- Composite reliability

## Validity test

- Convergent validity
- Discriminant validity

	$\alpha$	CR	AVE	CU	TD	FI	HE	IN	INTE	SA
CU	0.94	0.96	0.89	0.95 <sup>a</sup>						
TD	0.88	0.93	0.81	0.58	0.90					
FI	0.92	0.94	0.76	0.74	0.78	0.87				
HE	0.94	0.96	0.89	0.79	0.59	0.74	0.94			
IN	0.88	0.92	0.74	0.51	0.48	0.51	0.59	0.86		
INTE	0.88	0.92	0.80	0.61	0.54	0.60	0.66	0.70	0.89	
SA	0.88	0.93	0.80	0.41	0.38	0.44	0.56	0.59	0.59	0.90

**Notes:**  $\alpha$ , Cronbach's  $\alpha$ ; CR, composite reliability. CU, CUriosity; HE, Heightened Enjoyment; TD: Temporal Dissociation; FI: Focused Immersion; IN: Interactivity; INTE: INTEreat; SA: Satisfaction.

<sup>a</sup>The square root of AVE

Source: A. Hou, W. Shiau, & R. Shang (2019). *The involvement paradox. The role of cognitive absorption in mobile instant messaging user satisfaction.*

# Reliability and validity test

- Composite reliability:  $\alpha > 0.70$
- Convergent validity: AVE (avevar)  $> 0.50$
- Discriminant validity:  $\omega > 0.7$

```
library(semTools)
reliability(cfa_fit) %>% round(3)
```

	SI	JS	AC	EP	OC
alpha	0.886	0.281	0.891	0.847	0.823
omega	0.887	0.640	0.893	0.850	0.827
omega2	0.887	0.640	0.893	0.850	0.827
omega3	0.887	0.641	0.893	0.850	0.818
avevar	0.664	0.535	0.677	0.587	0.552

**Let's practice**

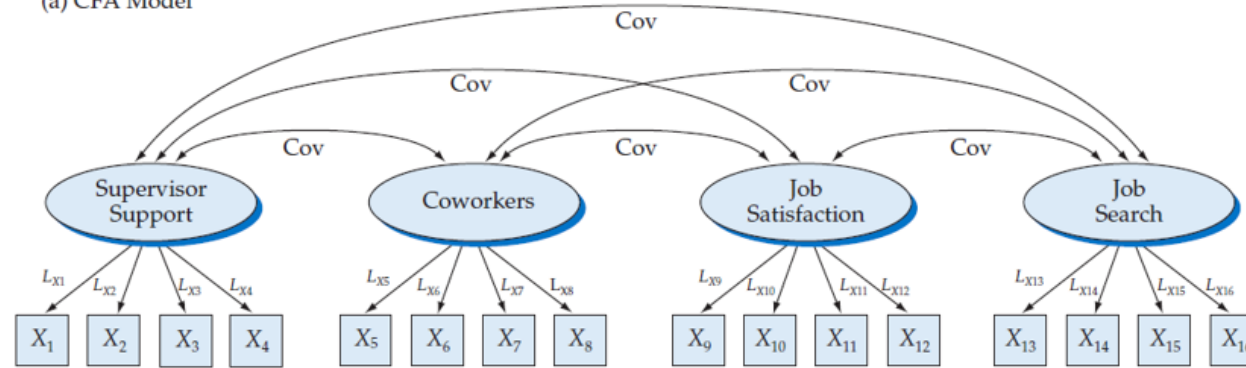
## Step 4: Specifying the structural model

---

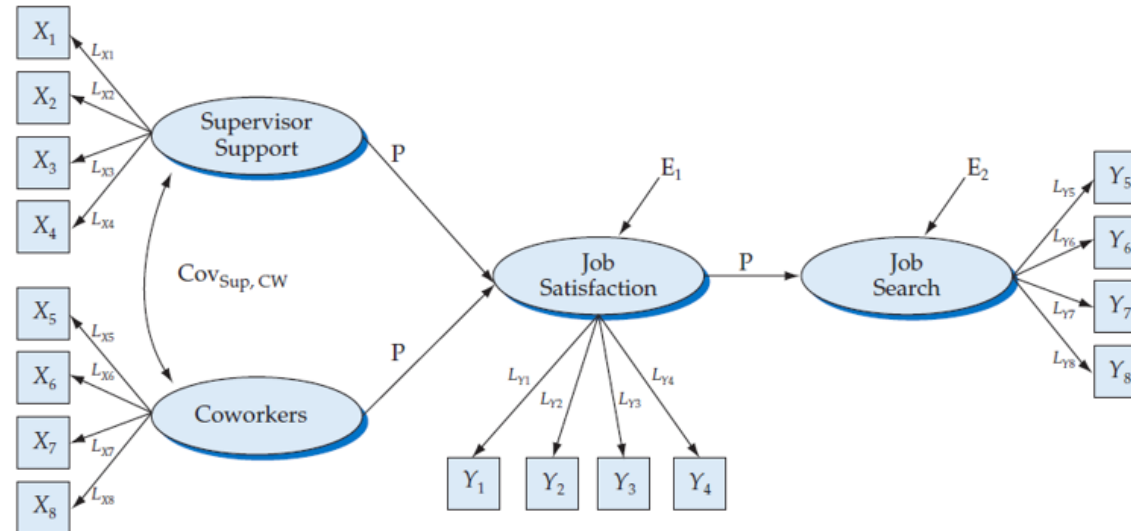


# CFA model to structural model

(a) CFA Model



(b) Structural Model

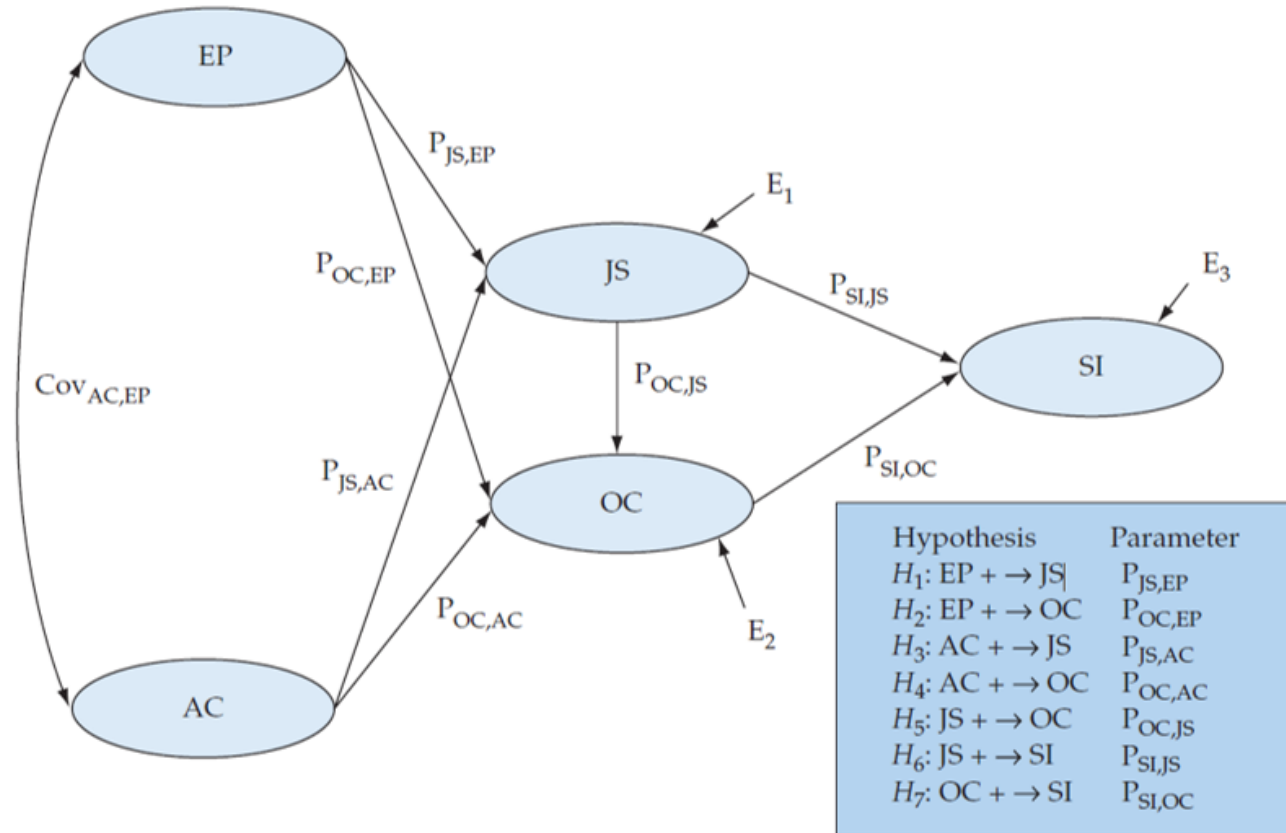


# Defining structural model

## Hypothesis:

- H1: Environmental perceptions are positively related to job satisfaction.
- H2: Environmental perceptions are positively related to organizational commitment.
- H3: Attitudes toward coworkers are positively related to job satisfaction.
- H4: Attitudes toward coworkers are positively related to organizational commitment.
- H5: Job satisfaction is related positively to organizational commitment.
- H6: Job satisfaction is related positively to staying intentions.
- H7: Organizational commitment is related positively to staying intention.

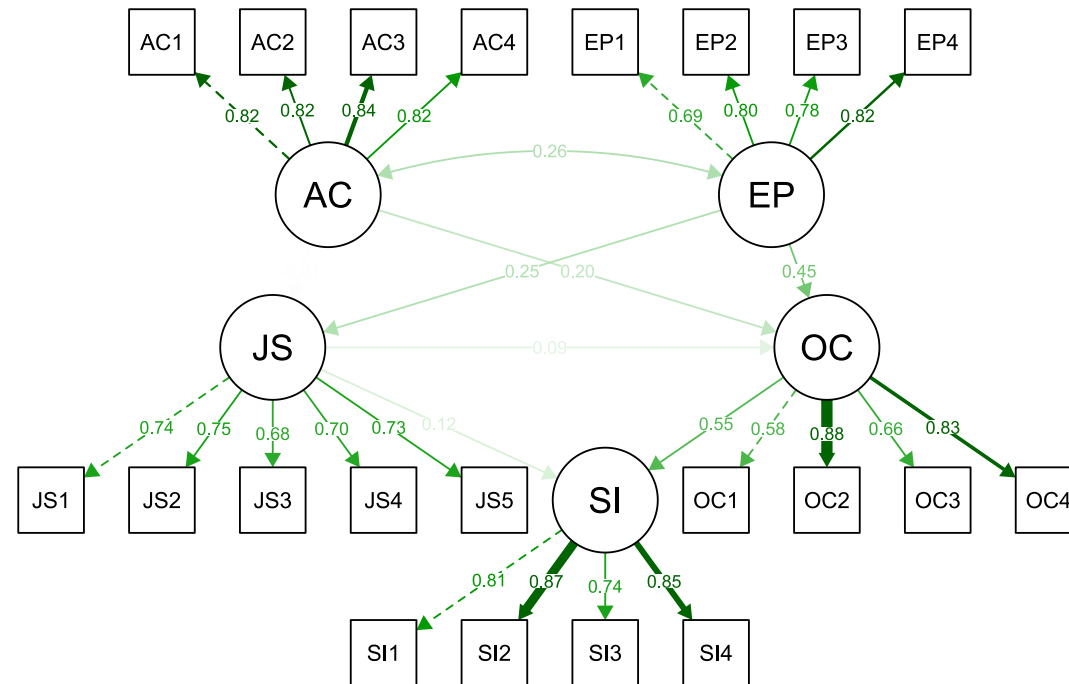
# Defining structural model



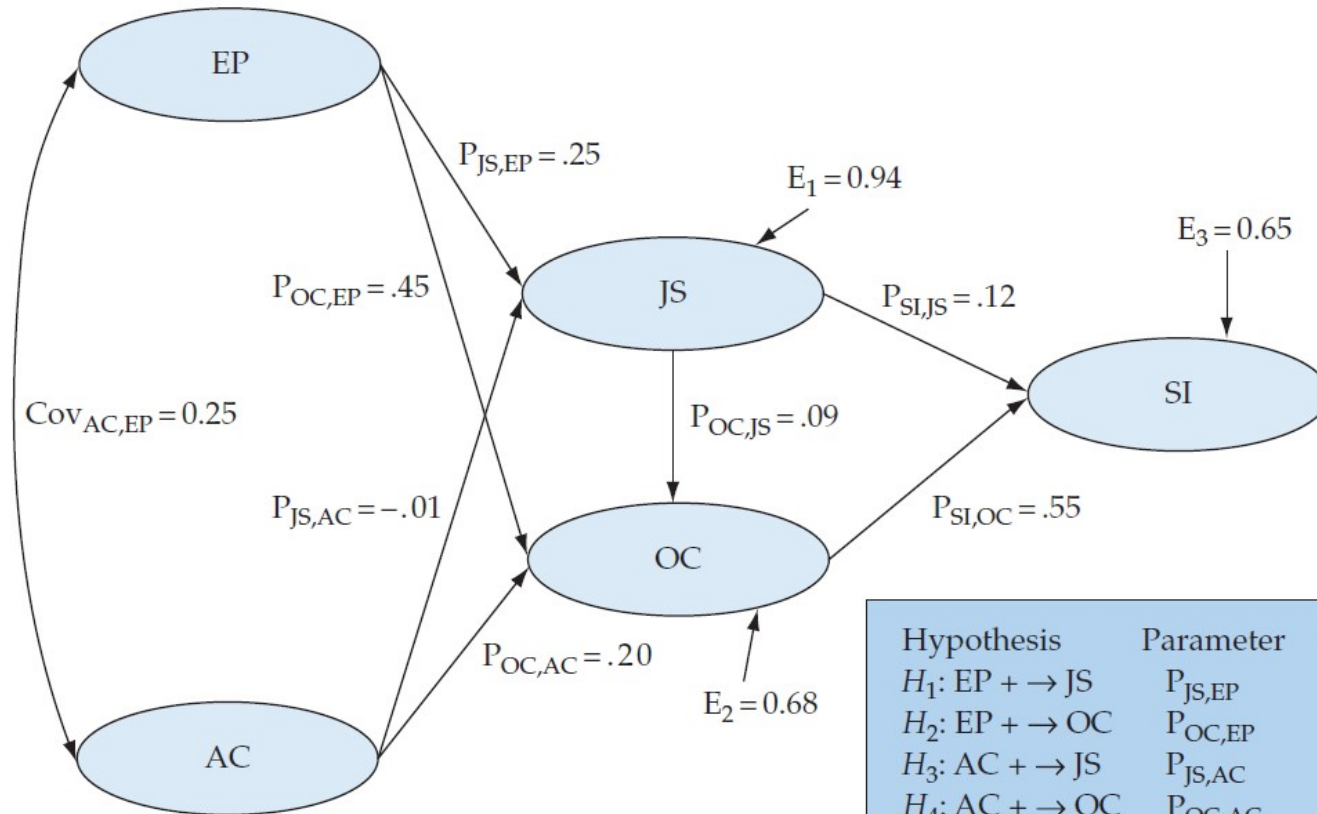
**Let's practice**

# Defining structural model

```
library(semPlot)
semPaths(object = sem_fit,
  what = "std",
  layout = "tree2",
  intercepts = FALSE,
  residuals = FALSE)
```



# Defining structural model



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

# GOF measures between structural and CFA model

```
gof_indices <- c('chisq', 'df','pvalue',  
                'rmsea', 'rmr', 'srmr',  
                'nnfi', 'cfi', 'agfi')  
fitmeasures(sem_fit, fit.measures = gof_indices)  
fitmeasures(cfa_fit, fit.measures = gof_indices)
```

chisq	df	pvalue	gfi	rmsea
287.179	181.000	0.000	0.938	0.038
agfi				
0.921				

chisq	df	pvalue	gfi	rmsea
240.738	179.000	0.001	0.947	0.029
agfi				
0.932				

GOF index	Employee retention model	CFA model
$\chi^2$ (chi-square)	287.179	240.738
Degrees of freedom	181	179
Probability	0.000	0.001
GFI	0.938	0.947
RMSEA	0.038	0.029
RMR	0.410	0.414
SRMR	0.060	0.036
NFI	0.936	0.946
NNFI	0.971	0.983
CFI	0.975	0.985
AGFI	0.921	0.932

# What's next?

- Modification indices
- Handling heywood cases
- Comparing competing models
- Formative scales in SEM
- Higher-order factor analysis
- Multigroup analysis





# Thank you!

Slides created via the R packages:



xaringan by Yihui



xaringanthemer and xaringanExtra  
by Garrick