

Introduction to Confirmatory Factor Analysis

A Quantitative Methods Consulting Center Workshop

Slides and materials available at
www.matthewmadison.com/workshop

Quantitative Methods Consulting Center (QMC)

- QMC purpose is to provide quantitative research support and training for members of the MFECOE
- QMC Team



Matthew Madison

Associate Professor
of Quantitative
Methodology,
QMC Director



Nancy Alila

Ph.D. Student in
Quantitative
Methodology,
QMC Research
Assistant



Ashley Heady

Ph.D. Student in
Quantitative
Methodology,
QMC Research
Assistant



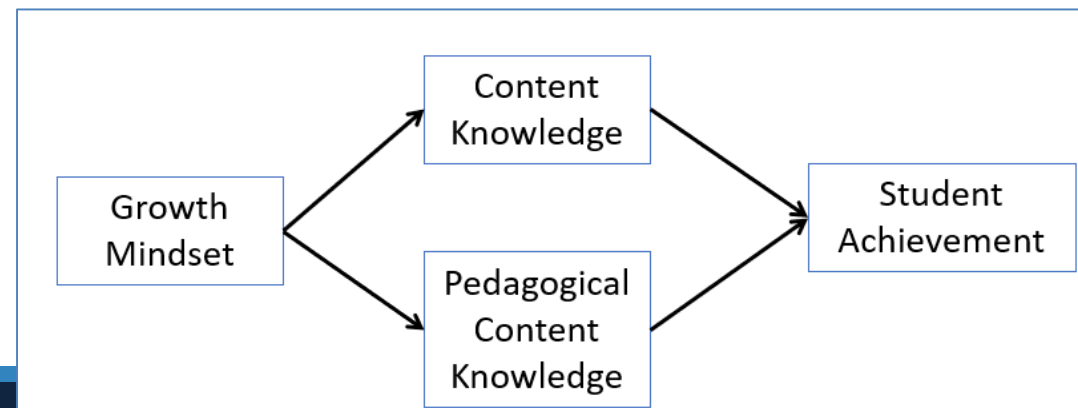
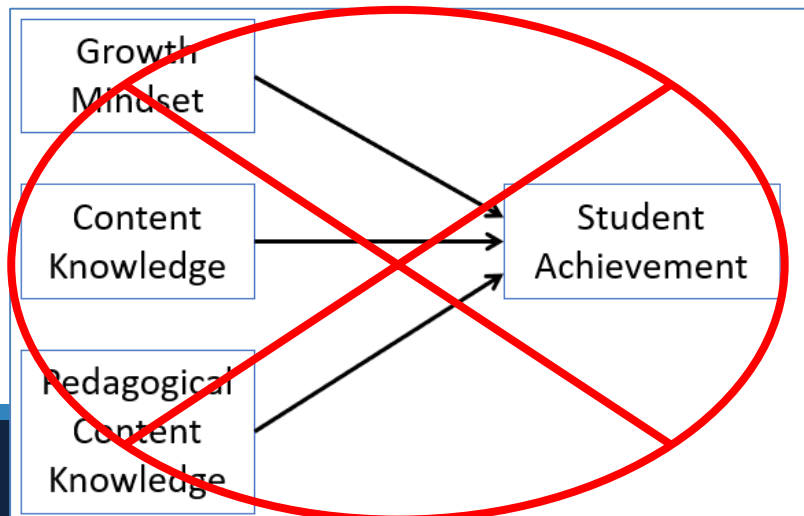
Beltran Pantoja

Ph.D. Student in
Quantitative
Methodology,
QMC Research
Assistant

QMC Fall Methods Workshop Series

- “Basics of Measurement Modeling”
 - Introduction to structural equation modeling (SEM)
 - Slides and materials available at matthewmadison.com/workshop
- SEM is comprised of two foundational methods

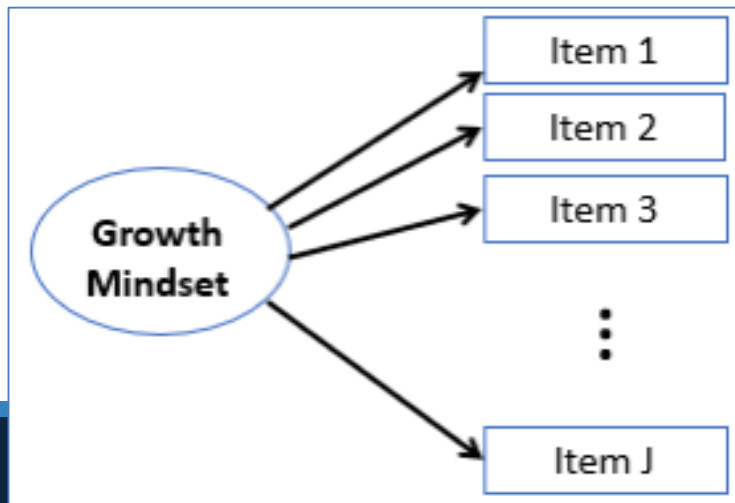
1. Path analysis



QMC Fall Methods Workshop Series

- “Basics of Measurement Modeling”
 - Introduction to structural equation modeling (SEM)
 - Slides and materials available at matthewmadison.com/workshop
- SEM is comprised of two foundational methods

2. Confirmatory Factor Analysis (CFA)

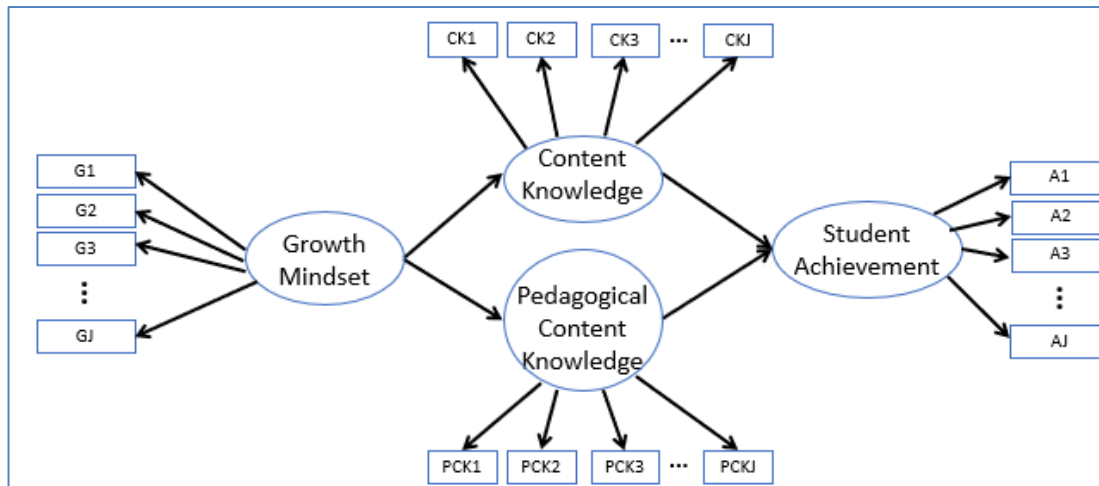


- **Benefits of CFA**
 - Properly model latent variables
 - Assess instrument reliability
 - Assess item quality

QMC Fall Methods Workshop Series

- “Basics of Measurement Modeling”
 - Introduction to structural equation modeling (SEM)

3. SEM = path analysis + CFA



- **Benefits of SEM**
 - Model complex relationships, test nuanced theories
 - Handle measurement error appropriately
 - Improved power and inferential accuracy
 - SEM is foundational for many other methods and techniques

Introduction to Confirmatory Factor Analysis

A Quantitative Methods Consulting Center Workshop

Slides and materials available at
www.matthewmadison.com/workshop

Agenda

- Conceptual introduction
 - Motivation
 - Technical foundation
 - Evaluating fit
- Empirical demonstration in Mplus
 - Evaluating and improving fit
- Resources and extras

Classical Test Theory

- Natural and intuitive
 - What you would naturally do
- Simply add each person's items to obtain an overall score
- In Classical Test Theory (CTT)
 - That overall score is called an **observed score (X)**
- In CTT, the overall score X has two components:
 - $X = T + E$
 - T is the true score; E is error

Reliability

- In short, reliability is related to the consistency of a score
- Different conceptualizations:
 - Multiple measurement occasions
 - Proportion of variance in observed score due to true score
- Why does reliability matter?
 - Affects our inferences

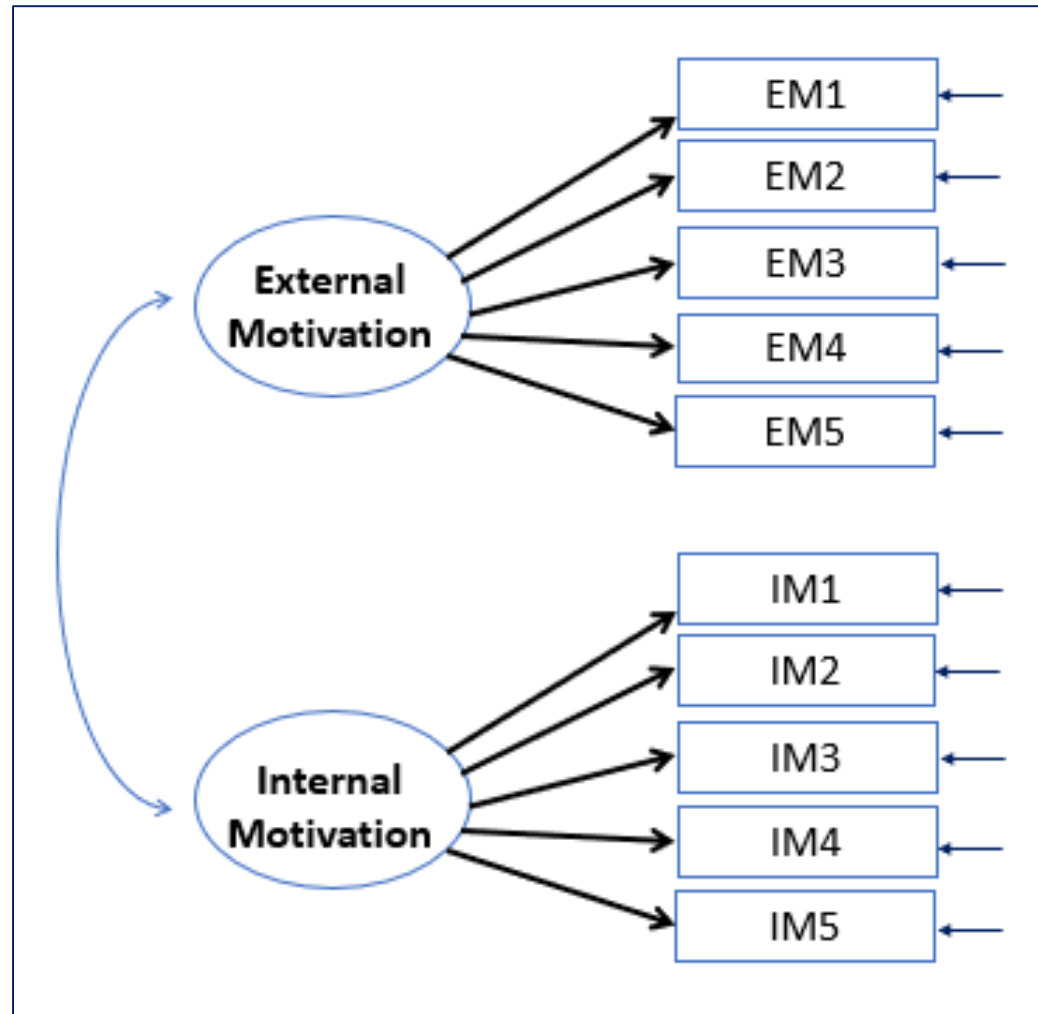
Measurement Models and CFA

- Difference between path analysis and a full structural model is that path analysis includes only observed variables
 - SEM includes latent variables
- Latent variables have measurement error
 - Observed variables (e.g., height, income) do not
 - Path coefficients in SEM are corrected for measurement error
 - Path analysis cannot correct for unreliability, therefore estimates are biased
- CFA accounts for measurement error and unreliability in SEM

CFA Models

- **CFA are based on theory**
- CFA offers a method for the quantification and testing of theories
- Can specify a model attached to substantive considerations that represent plausible explanations of empirical data
 - # of latent factors and which observed variables represent constructs

CFA Models



CFA Models

- **CFA are based on theory**
- CFA offers a method for the quantification and testing of theories
- Can specify a model attached to substantive considerations that represent plausible explanations of empirical data
 - # of latent factors and which observed variables represent constructs
 - In above example, two factor: internal and external motivation
 - Items IM1-IM5 are theorized to be explained by the internal motivation factor
 - Items EM1-EM5 are theorized to be explained by the external motivation factor

EFA Models

- Exploratory factor analysis is used when there is no theory or hypothesized structure
- Run several models and compare fit
 - Determine the number of factors
 - How the factors load onto items
- Then follow-up with a CFA on a different sample (or subsample)

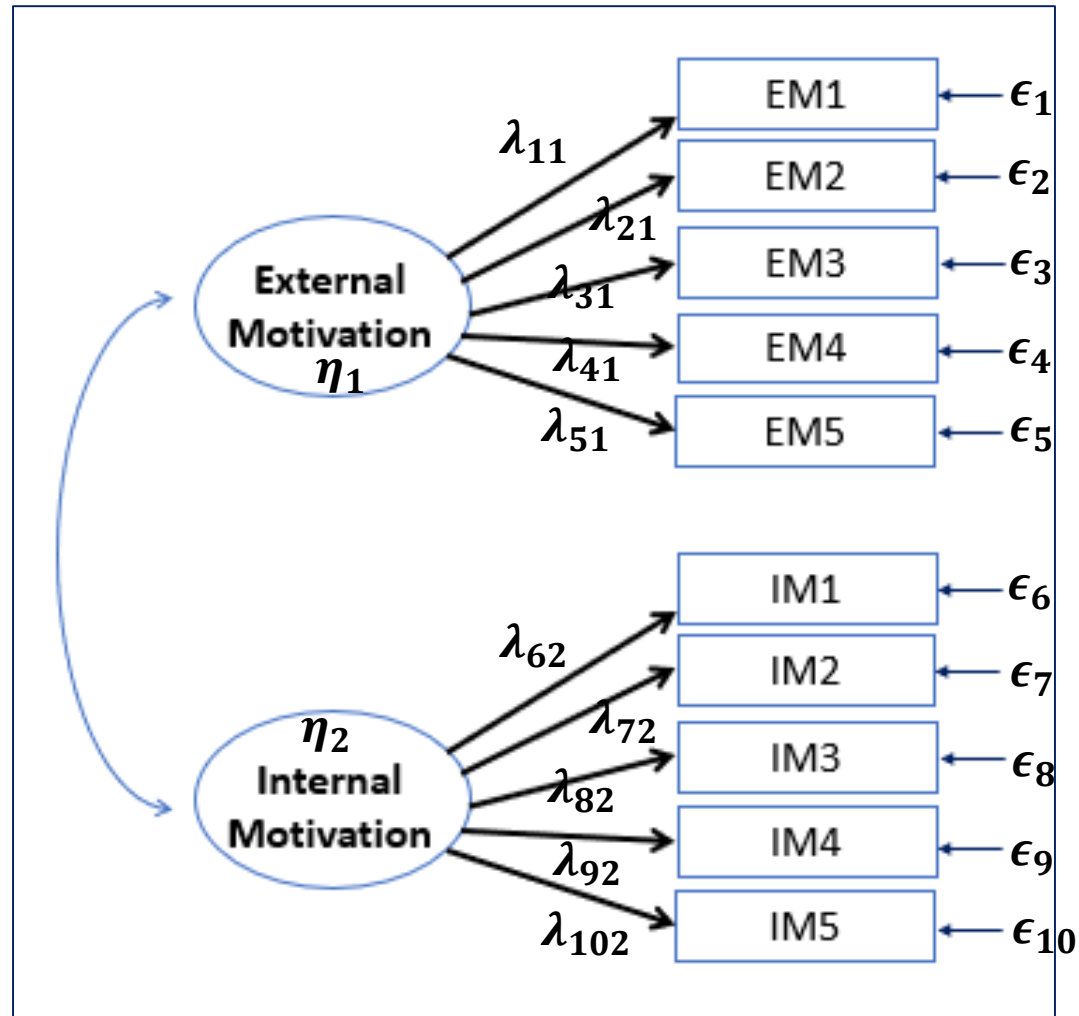
CFA Models

- Each indicator is *typically* hypothesized to be explained by one latent factor (simple structure)
- In the previous example:
 - Internal motivation doesn't have a direct effect on external motivation
 - External motivation doesn't have a direction effect on internal motivation
- One doesn't have to specify simple structure
 - Just makes interpretation easier

CFA Identification

- These are latent constructs; so the scale needs to be set
 - Either fix first loading to 1 (Mplus default), or
 - Fix factor variance to 1
- Must be a certain number of indicators (items)
 - One factor: at least 3 indicators
 - More than one factor: at least 2 per factor, preferably more to avoid empirical under-identification

CFA Equations



CFA Equations

- Mini-regression models for each item
- Generally: item score = intercept + loading*factor + error
-
- $em_1 = \lambda_{11}\eta_1 + \lambda_{12}\eta_2 + \epsilon_1 = \lambda_{11}\eta_1 + \epsilon_1$
- $em_2 = \lambda_{21}\eta_1 + \epsilon_1$
- ...
- $im_4 = \lambda_{91}\eta_1 + \lambda_{92}\eta_2 + \epsilon_1 = \lambda_{92}\eta_2 + \epsilon_9$
- $im_5 = \lambda_{102}\eta_2 + \epsilon_{10}$

CFA Equations

- Just like a regression, except the variables are latent
- Interpret like linear regression
- **Unstandardized** coefficient is unit change in indicator for every unit change in factor.
 - If there is more than one factor influencing indicator, the coefficient is unit change in indicator for every unit change in factor *controlling for other factors*
 - Relationships between latent factors are covariances
- **Standardized** coefficient equals correlation between indicator and factor if only 1 factor is directly affecting indicator
 - If more than one factor, interpret like a standardized regression coefficient (e.g., standard deviation units, holding other factors constant)

Model Fit Evaluation

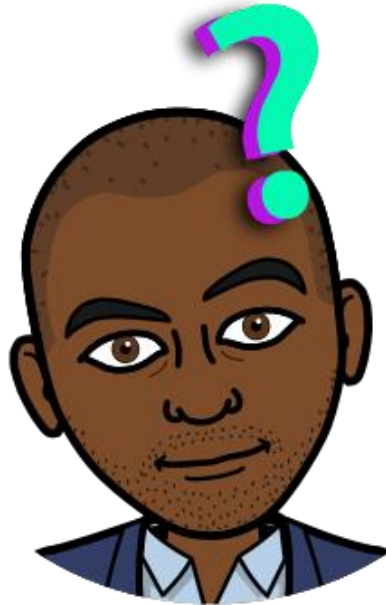
- Several tests and metrics should be used to evaluate fit
- Chi-square test of model fit
 - $p > .05$ indicates good fit, but very sensitive with large samples
- Root mean square error (RMSEA)
 - Less than .05 indicates good fit, or .08 for acceptable fit
- Standardized Root Mean Square Residual (SRMR)
 - Less than .05 indicates good fit, or .08 for acceptable fit
- Comparative/Tucker-Lewis Fit Indices (CFI/TLI)
 - Greater than .90 or .95 indicate good fit

Model Fit Improvement

- Examine loadings and factor correlations
- Examine standardized residuals (bigger than 2 or 3 are flags)
- Modification indices
 - Suggestions on model adjustments to improve fit
 - Caution: modification indices theory agnostic and rarely lead to true model
- All these statistics provide information that can help researchers better understand the relationships and lack of fit

Questions

- Next is empirical demonstration in Mplus



Empirical Example

- Native STEM Portraits
 - Mixed methods, longitudinal study of Native individuals in STEM majors and careers
 - Survey designed to measure important constructs
 - Had a hypothesized structure

CFA Resources

- Measurement Theory text (Bandalos, 2018)
- YouTube
 - CenterStat, QuantFish, QMC YouTube channel, many more
- Courses
 - ERSH 7600: construction of measurement instruments
 - ERSH 8130: survey of latent variable methods
 - ERSH 8750: structural equation modeling
 - ERSH 8740: exploratory factor analysis

CFA Applications

- Lots more can be done in a CFA
- Measurement invariance testing
 - Instrument function similarly across subgroups of interest?
 - If not, why?
- Latent mean testing

More Questions

- Visit us in the Quantitative Methods Consulting Center
- tinyurl.com/uga-qmc
- Aderhold 321A, qmcenter@uga.edu

