CFA

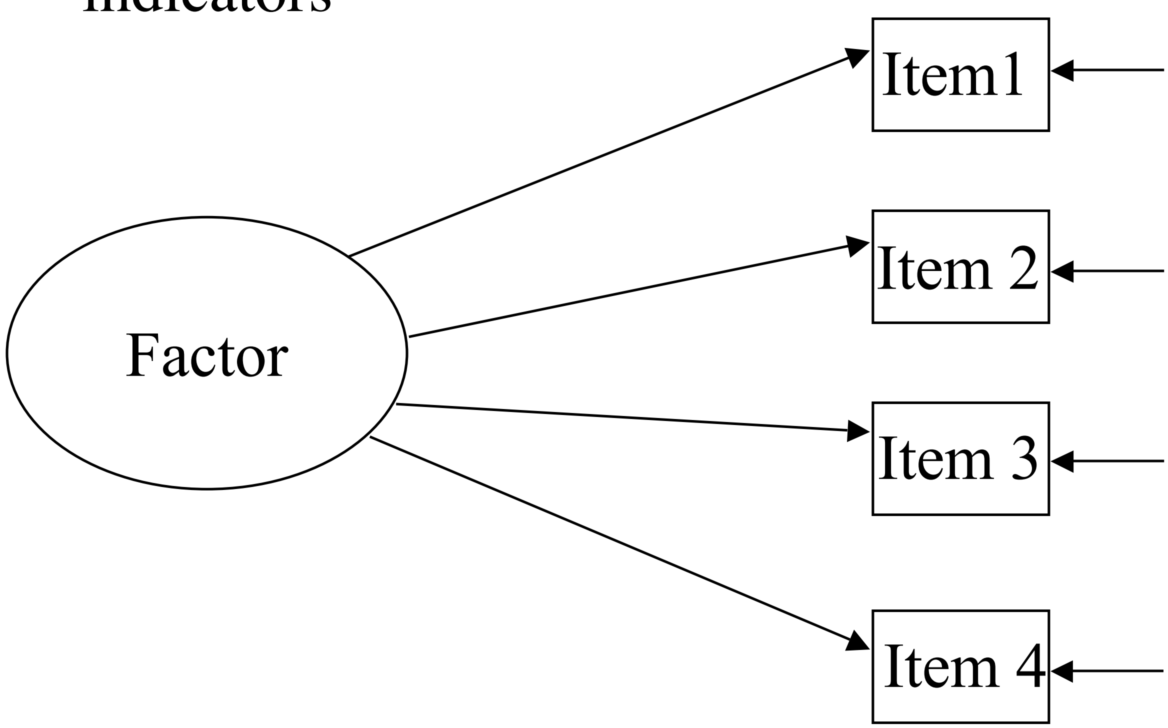
- EFA vs. CFA
- Review of Conceptual Model
- Cause vs. Effect Indicators
- Specification
- Identification
- Setting the scale of latent variables
- Testing Models
- Hierarchical Models
- Russell
- Worthington
- Some examples

Confirmatory Factor Analysis

- EFA: Data exploration
  - » find model that best fits data
- CFA: Hypothesis testing
  - » define a model and test whether the data support it
- Many testable hypotheses
  - » number of factors
  - » orthogonality of factors
  - » relationship of factors to items (which items load on which factors)
  - » size of loadings
  - » cross-group equivalence
  - » and more

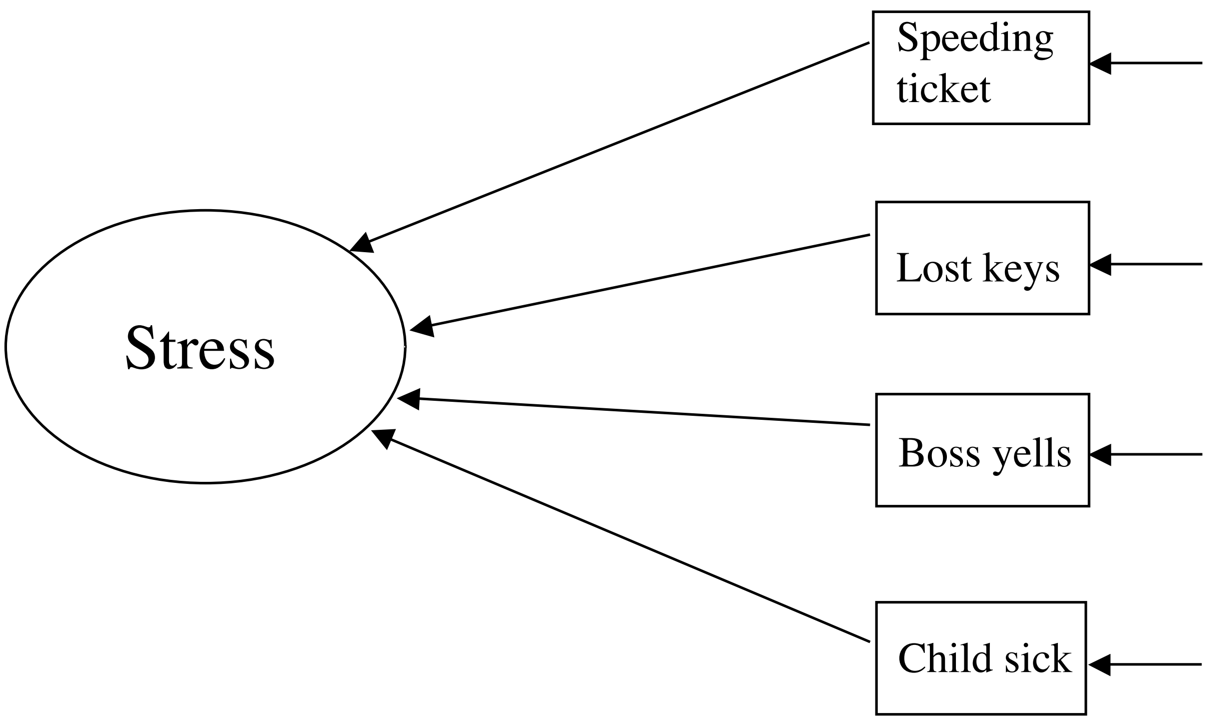
Appropriateness of the Model

- FA assumes a particular causal model
  - » Unmeasurable factors cause measured variables
  - » All variance shared between variables is due to the factors they have in common
  - » called "effect indicators" or "reflective indicators"



A Different Model

- A different model may be appropriate
- E.g., sometimes it's more plausible that the causal flow is in the opposite direction
  - » daily hassles
- Called "cause indicators" or "formative indicators"

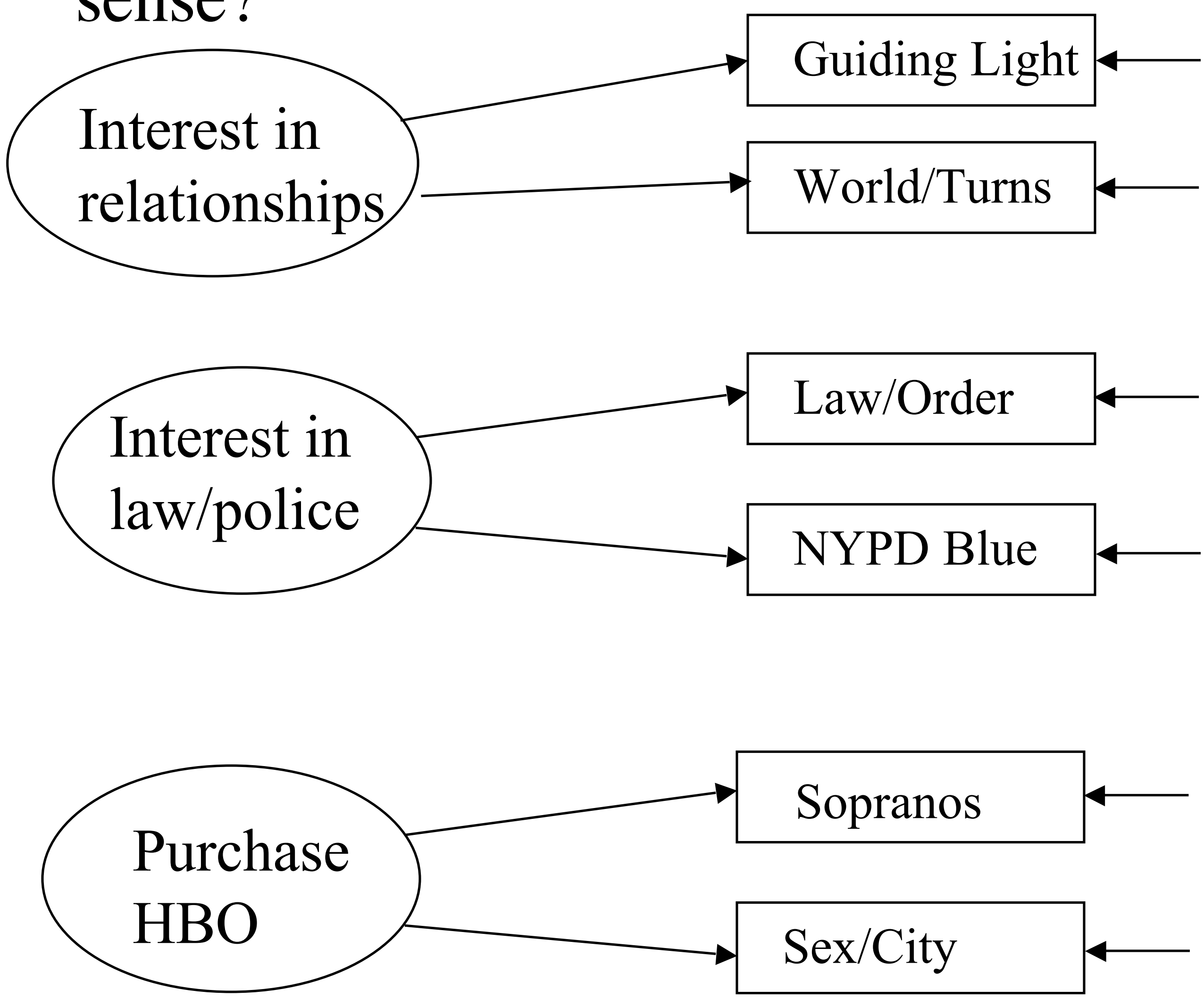


Another Example

- Research question: Effect of television sex and/or violence on behavior
- Participants report frequency of viewing several dozen programs
- Factor analysis results:
  - » F1: Law/Order, NYPD Blue, Practice, Third Watch
  - » F2: Jerry Springer, Simpsons, South Park
  - » F3: 7th Heaven, Dawson's Creek, Felicity
  - » F4: ER, Judging Amy, Providence, Touched/Angel, West Wing
  - » F5: All My Children, General Hospital, One Life to Live
  - » F6: As World Turns, Young/Restless, Guiding Light
  - » F7: Sex and the City, Sopranos

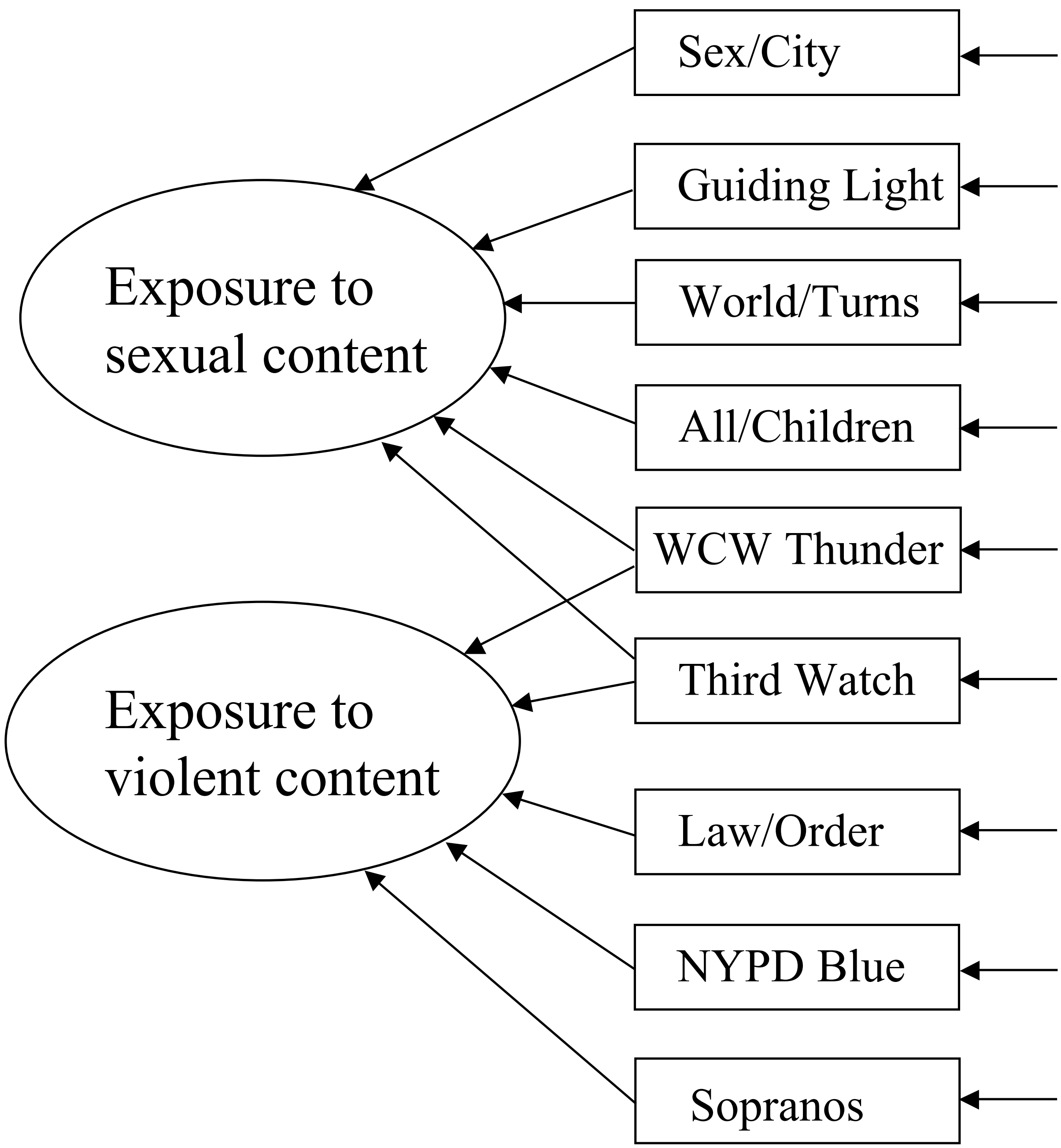
Common Factor Model of TV Viewing

- Does the common factor model make sense?



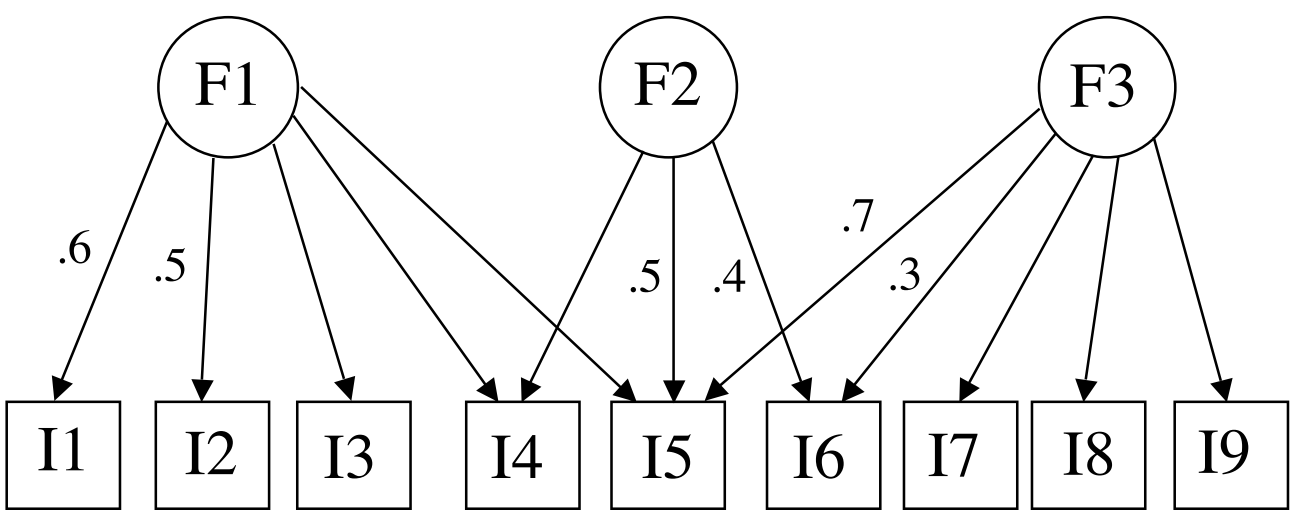
Alternative Model of TV Viewing

- Better fit with our hypotheses/interests



How CFA Works

- Research specifies a model
- Software computes loadings
- Model + loadings imply a set of inter-item correlations
- Computed correlations are compared with actual (observed) correlations
  - » i.e., residuals are computed
  - » Small residuals = adequate (plausible) model
  - » Big residuals disconfirm the model



CFA Software

- In general CFA requires specialized software (e.g., LISREL, EQS, AMOS)
- Some very simple hypotheses can be tested with SPSS, using ML or GLS extraction method
  - » E.g., number of factors -- Is one factor adequate to explain the data? Does a three-factor model fit significantly better than a two-factor model?
- Note that if more than one factors is hypothesized, you can't specify which items load on which factors
  - » but, we can do a crude (ballpark) kind of theory confirmation

Specification Issues

- Specification of CFA models involves answering 3 questions

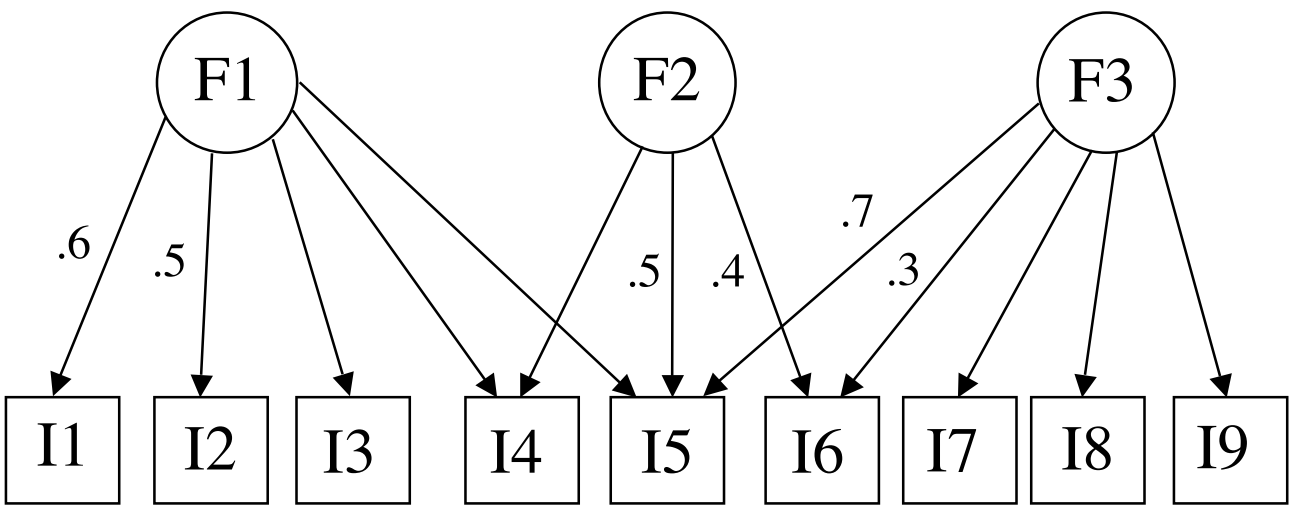
  » How many factors are present?

  » Which indicators are influenced by which factors?

  » If more than one factor, how are factors interrelated?

Specification

- How many factors?
- Which indicators are influenced by which factors?
- How are the factors intercorrelated?



Model Identification

- A model is identified if it is theoretically possible to derive a unique estimate for each parameter.
- For CFA models, two necessary (but insufficient) conditions in order to be identified.
- First, number of parameters must be less than or equal to the number of observations
- Second, each latent variable must have a scale

Sufficient conditions for identification

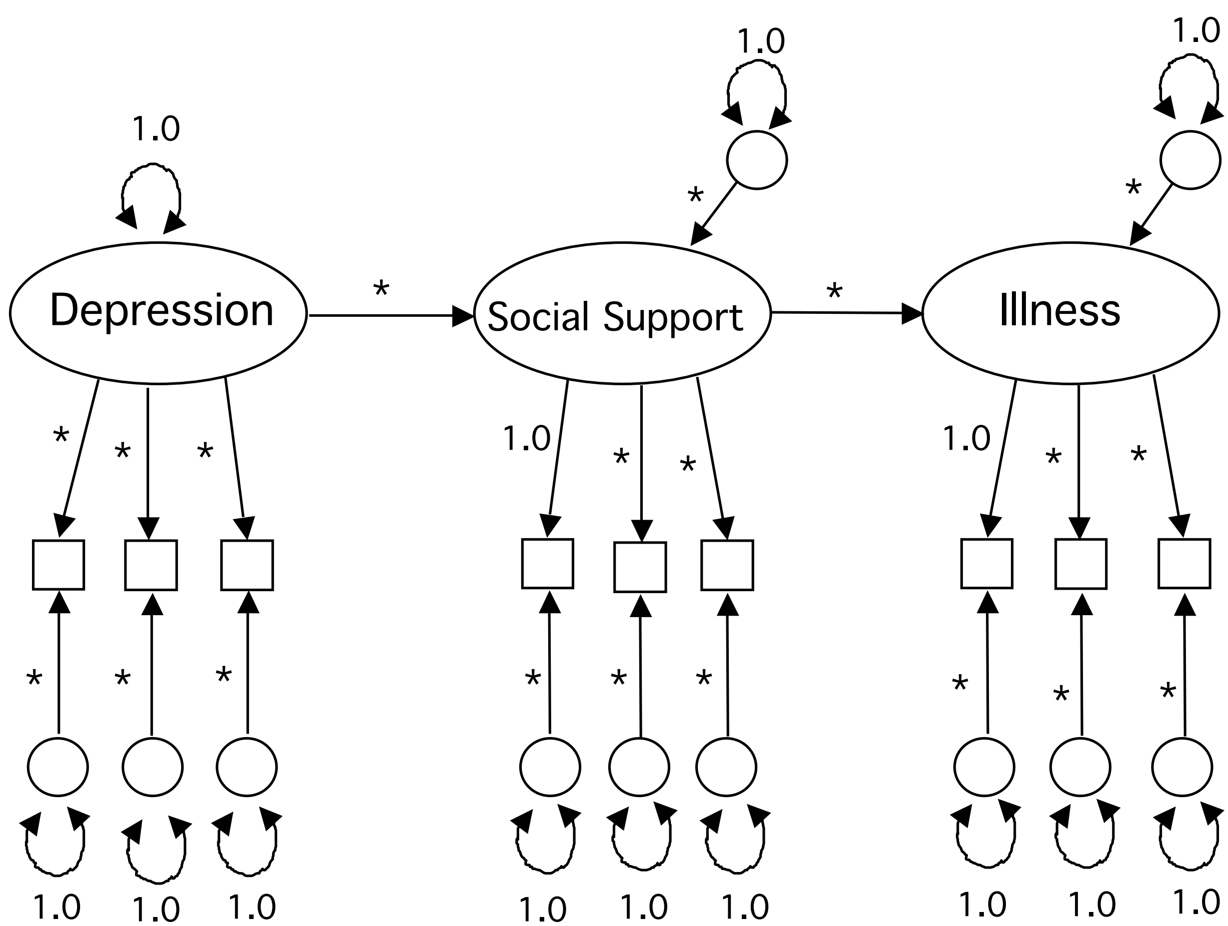
- Standard CFA model =
  - » unidimensional measurement (i.e., each indicator loads on one and only one factor)
  - » all possible unanalyzed associations between factors are included in the model (i.e., all factors are allowed to correlate)
- For standard CFA model with 1 factor
  - » must have three indicators for identification
- For standard CFA model with 2 or more factors
  - » each factor must have at least 2 indicators for identification
- For non-standard CFA models (e.g., with multidimensional measurement) no simple rules
  - » see Kline for empirical tests

Setting the Scale (of latent vars)

- Latent variables are not measured and thus are not tied to any concrete scale
  - » e.g., the RWA measure has a scale (so a score of 17 means something concrete)
  - » but the abstract construct "authoritarianism" does not have a scale (so we don't know what a score of 17 would mean)
- We must “set the scale” for each LV
- Two ways to do this

Two Methods for Setting Scale

- Fix the variance of the latent variable (typically to 1.0)
- Fix the value of one path leading from the latent variable to an indicator (the "reference variable") (typically to 1.0)



Setting the Scale

- Both methods generally result in same overall fit
  - » so choice is somewhat arbitrary
  - » but see Kline for a discussion of specialized cases in which this might not be true
- Setting the variance of a factor to 1.0 standardizes that factor
  - » probably not appropriate if comparing separate groups
  - » most software packages won't allow this if a factor is an endogenous variable (as hierarchical CFA models)
- Most commonly in CFA, we would set the variance to 1.0
- If setting a path coefficient, choose the indicator that is the most reliable

Automatic vs. Manual

- For disturbance terms, EQS automatically sets the scale for you
  - » by fixing the path coefficient to 1.0
- For other latent variables you must set the scale yourself
- If you forget, you will get error messages

PARAMETER	CONDITION CODE
F2,F1	LINEARLY DEPENDENT ON OTHER PARAMETERS
F2,F2	LINEARLY DEPENDENT ON OTHER PARAMETERS
V273,F3	LINEARLY DEPENDENT ON OTHER PARAMETERS

Setting the Scale - How To

- There are a variety of ways to do this
- If you are working with a diagram
  - » double-click on path you want to set OR
  - » double-click on factor whose variance you want to set
- Change parameter type from "free parameter" to "fixed parameter"
- Set start value to 1.0 (or whatever you want to fix the parameter to)
- Any fixed path is red in diagram
- Syntax changes:
  - » V260 = 1F1 + E260; (fixed)
  - » V261 = *F1 + E261; (free)
  - » F2 = *; (free)
  - » F3 = 1; (fixed)
  - » V261 = 2.00*F1 + E261; (free with a start value)

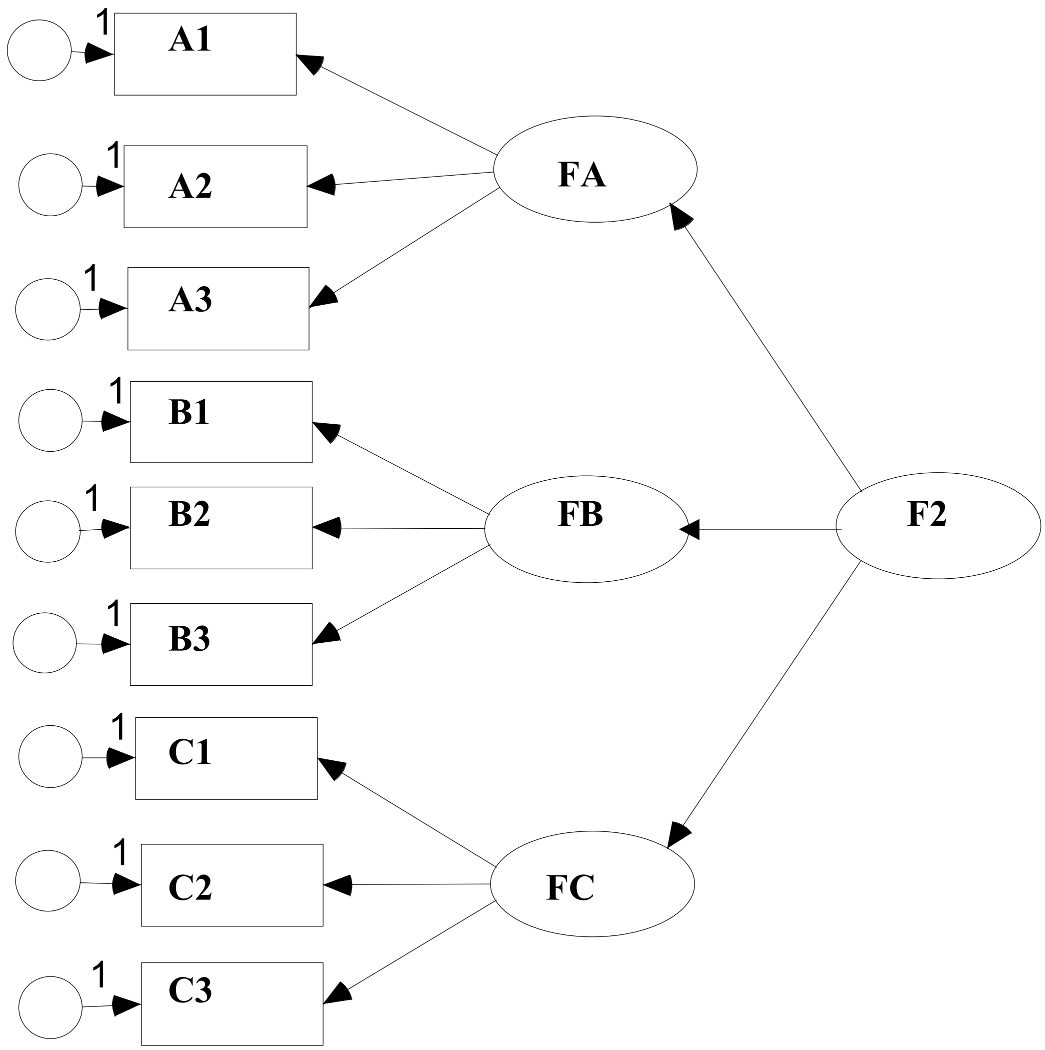
Testing CFA Models

- Often, you will have a specific theoretical model (or models) to test
- If not, Kline suggests a series of analyses
  - » first test a single-factor model
  - » if that model is poorly fitting (and only then) move on to test multiple factor models
  - » Can also do specific respecifications (adding or deleting paths, including correlated disturbance terms)

Second order factor models

» Just as latent variables might explain correlation among items, second order latent variables might explain correlation among factors

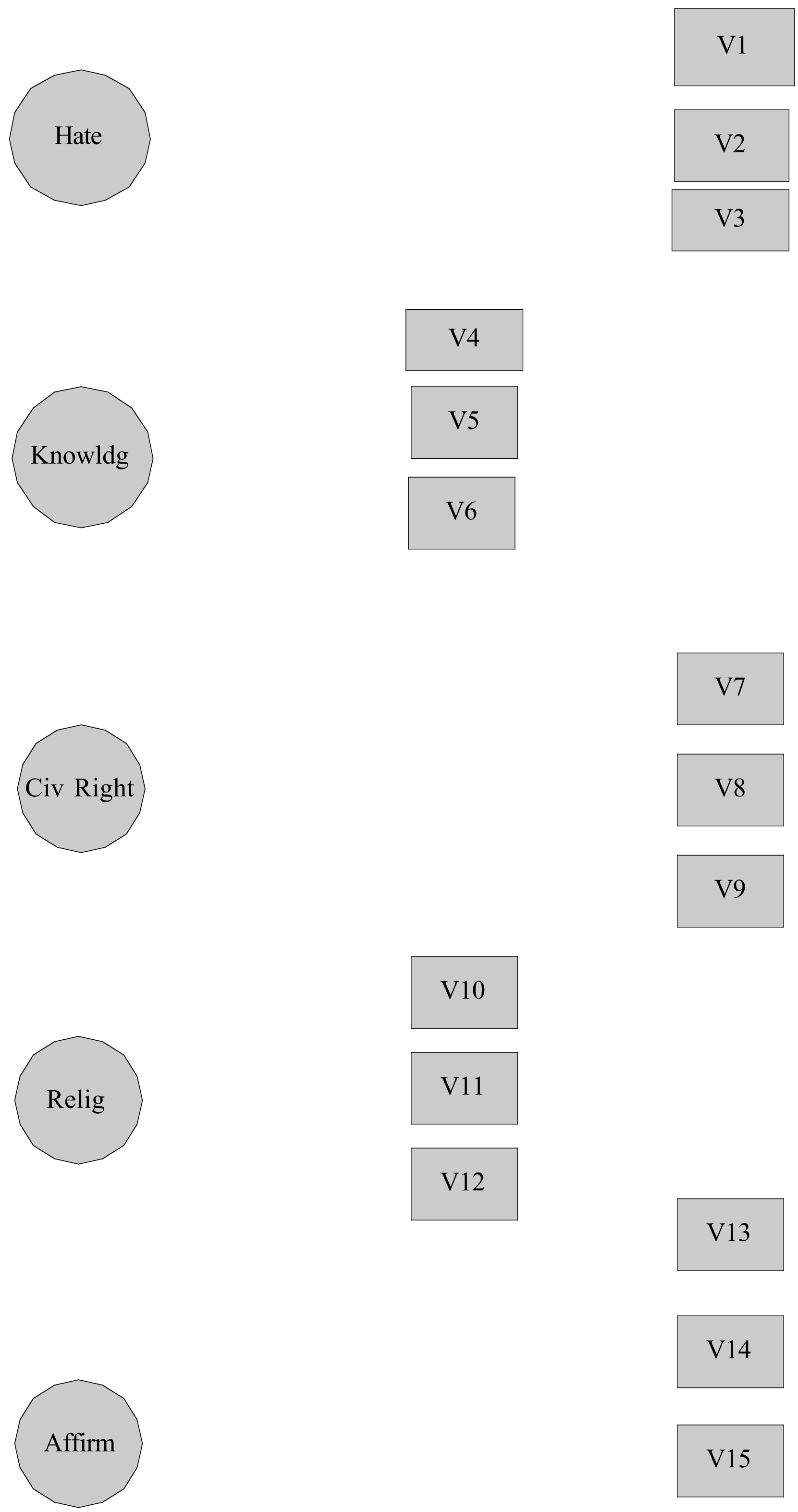
- 1st order factors now endogenous



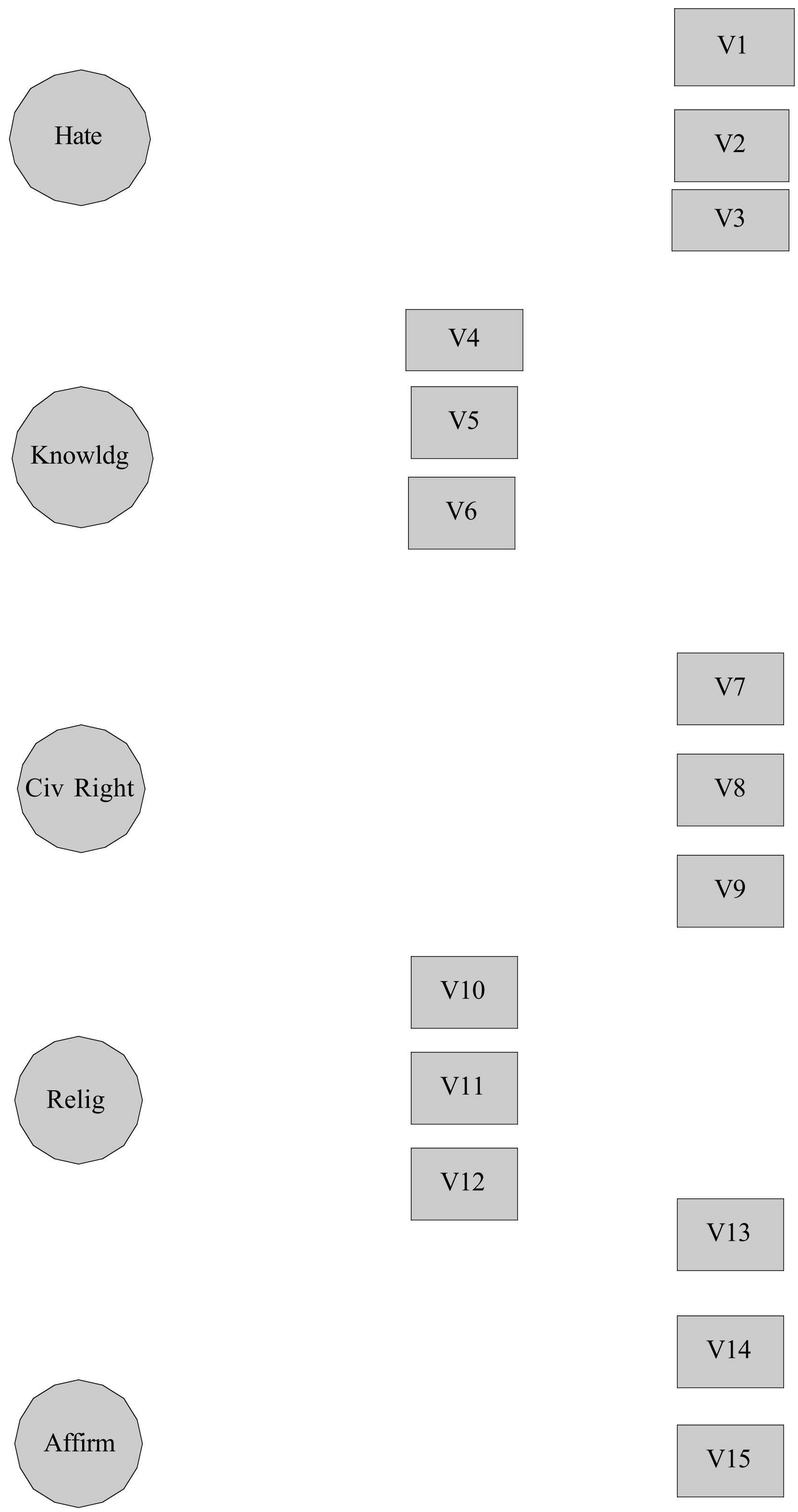
Comments on Russell (2002)

- Reviewed CFAs in PSPB
  - » half tested a hypothesized factor model
  - » half were testing the measurement part of a full (hybrid) SEM
- Reports new guidelines calling for more stringent standards for fit indices
  - » e.g. > .95 for incremental indices
  - » more variables makes this standard harder to achieve
  - » "may need to reduce the number of items on a scale"
- Recommends sample sizes of at least 100
  - » unless loadings are very high (around .80)
- Advanced techniques for dealing with non-normality or missing data are being developed

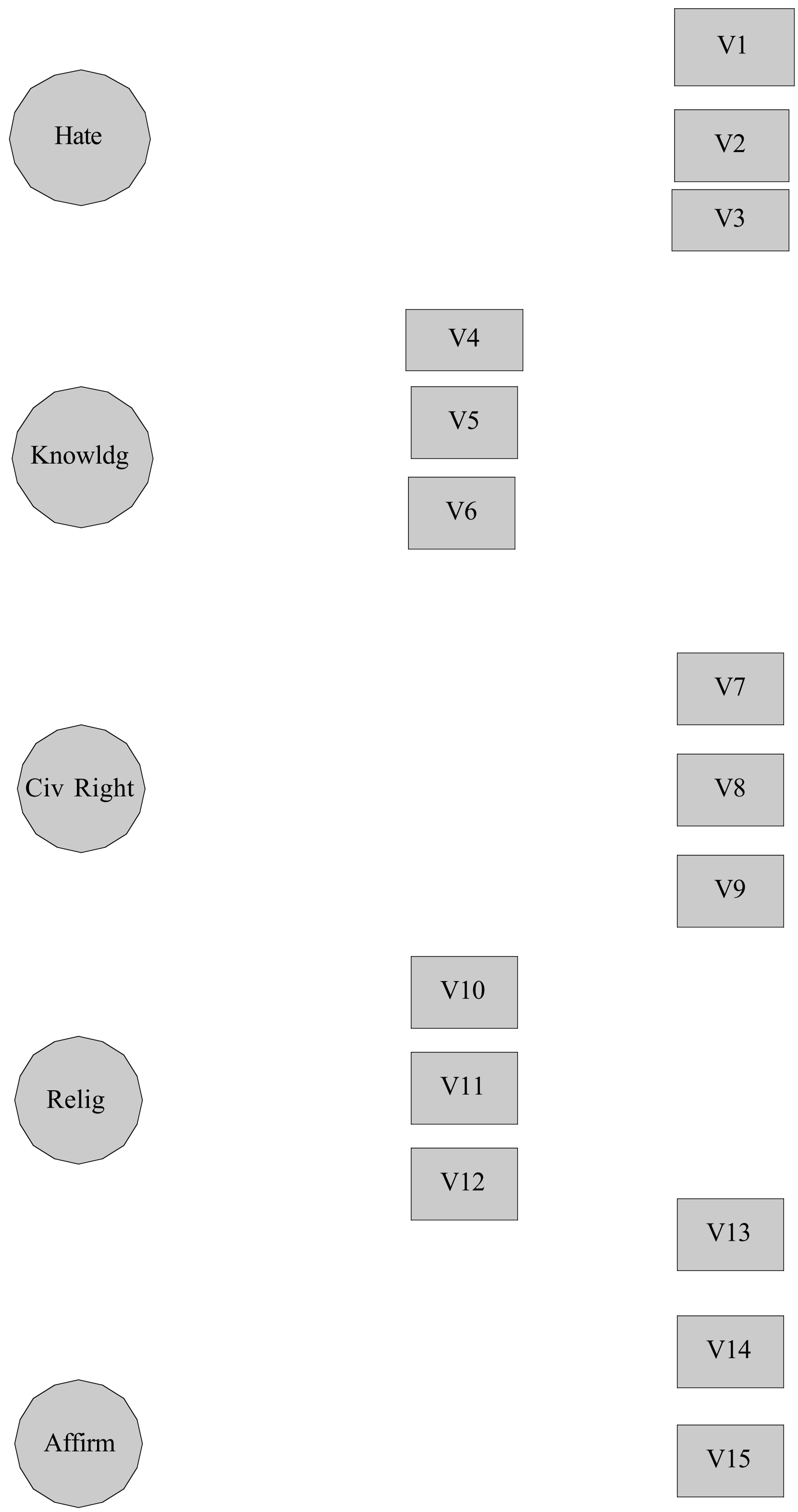
Worthington: Independence



Worthington: Five Factor Oblique



Worthington: Second Order

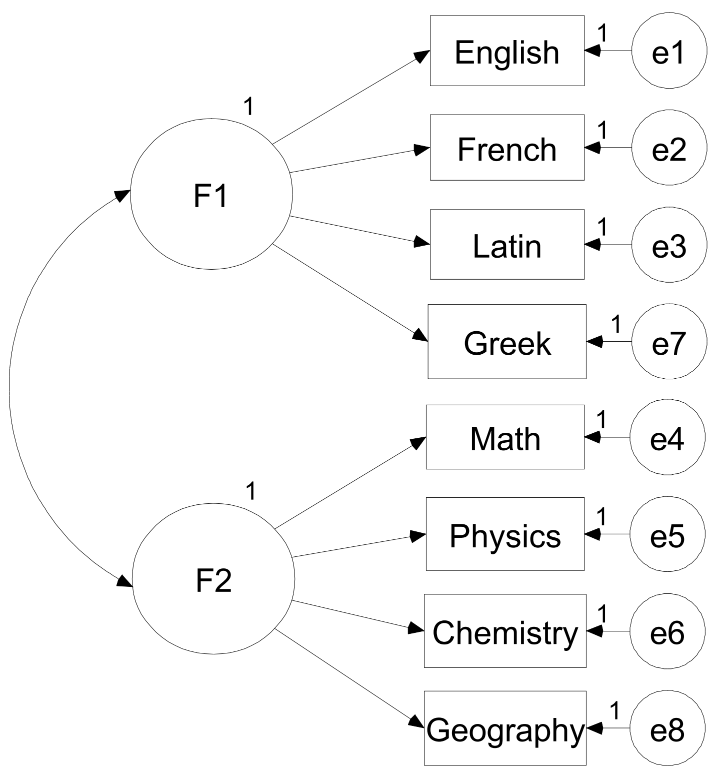


Questions on Worthington

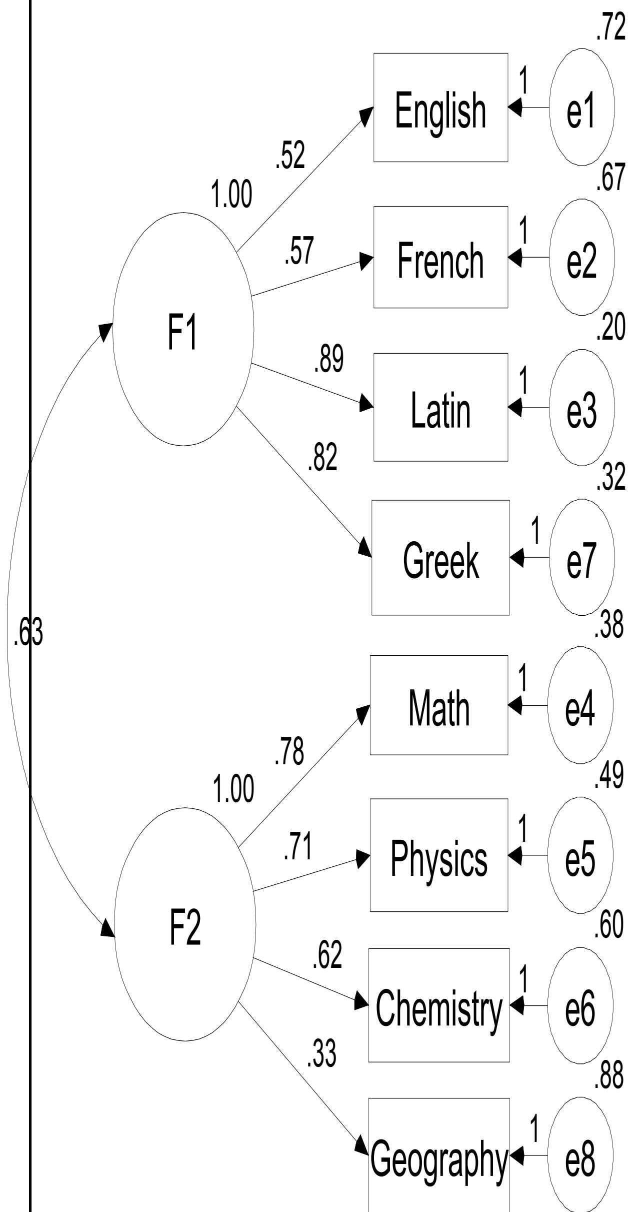
- "second-order model does not improve on the first-order model" (p. 112)
  - » is it possible that it could?
- "the IFI, PCFI, and RMSEA reached values that indicate reasonable fit" (p. 112)
  - » do you agree?
- "we were unable to locate any meaningful modifications in the model that might result in inceased fit. Therefore, post hoc respecification of the model was not attempted." (p. 113)
  - » do you think this was a good decision?

One-Factor vs. Two-Factor

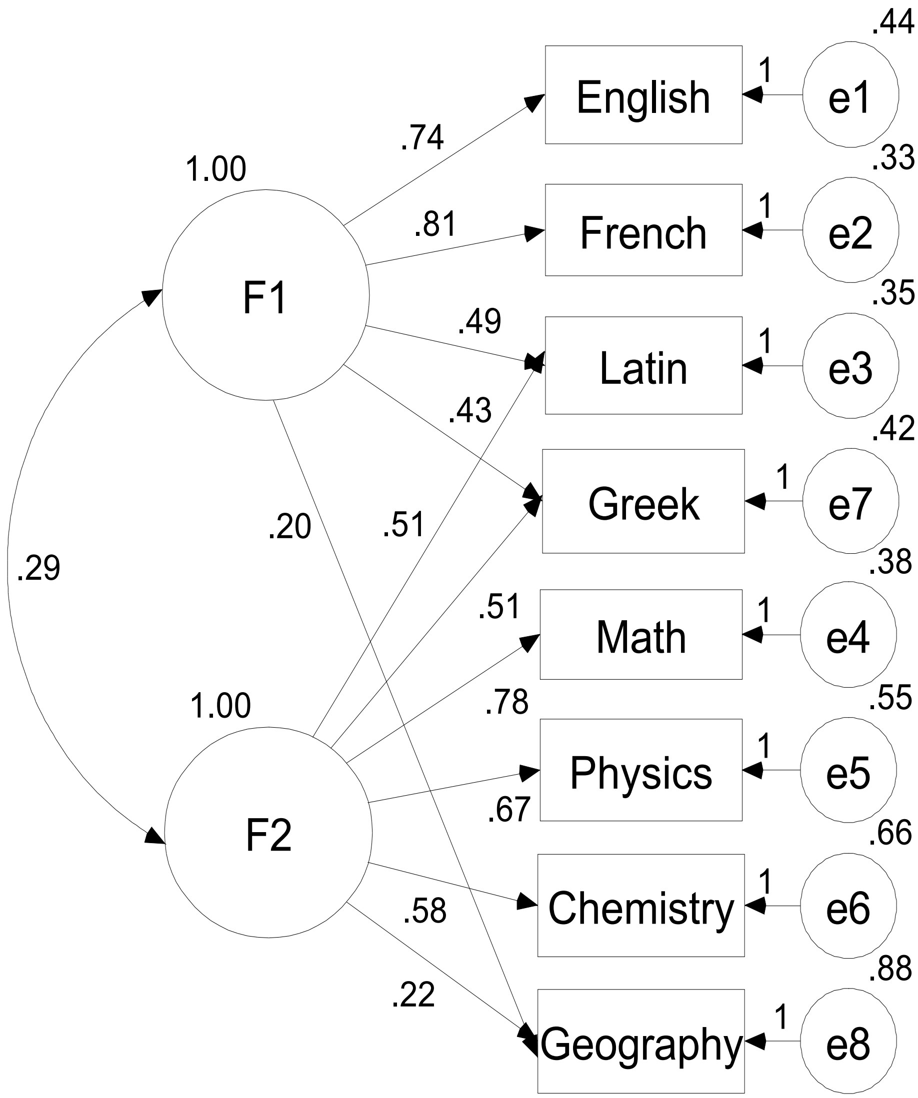
- Consider this two-factor model of knowledge of academic subjects
- One-factor model is nested within it
  - » don't you have to add and delete paths to get from one model to the other?
  - » no, constrain the correlation to 1.0 to get the 1-factor model
  - » can do chi-square difference test



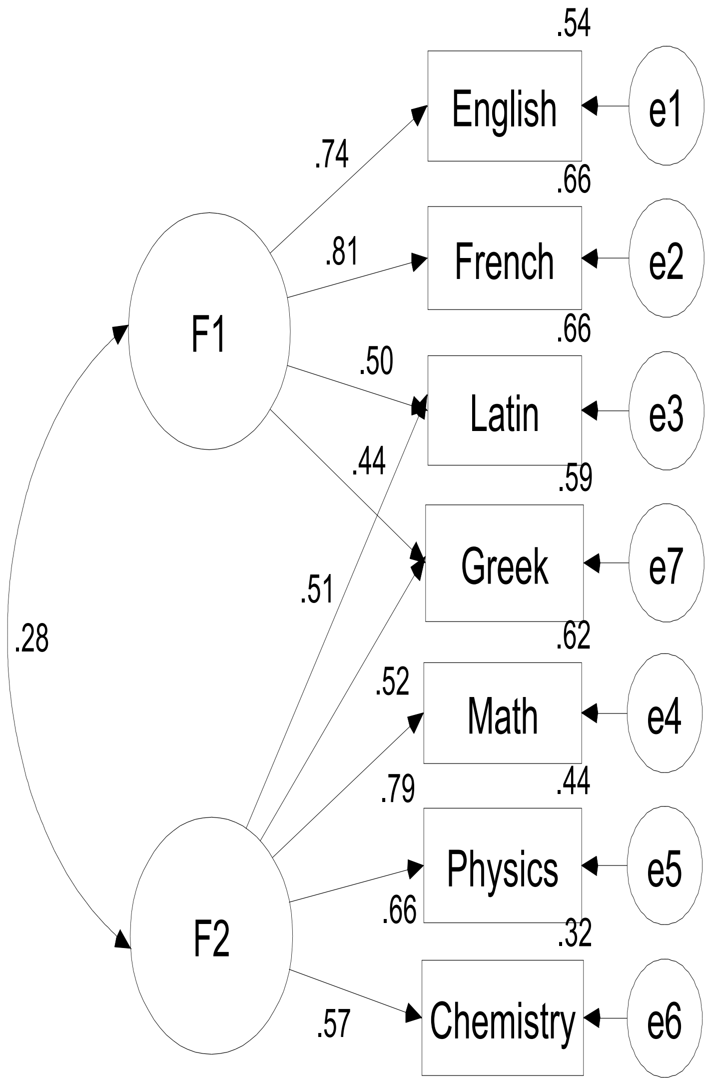
Two Ways to Set Scale



CFA 1a - Multidimensionality



CFA 1b



2-Factor Solution with Latin and Greek loading on both factors and Geography removed.

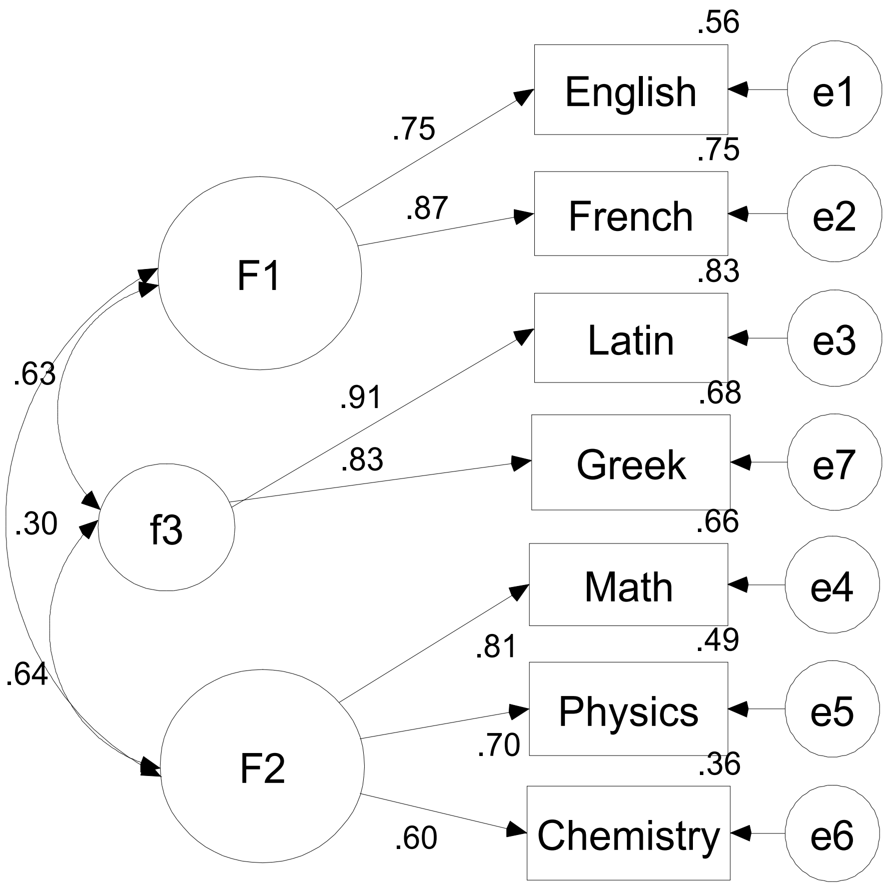
parameters estimated: 17

ChiSquare(11, N=150)=45.438, p=.000

RMSEA=.145, RMR=.050

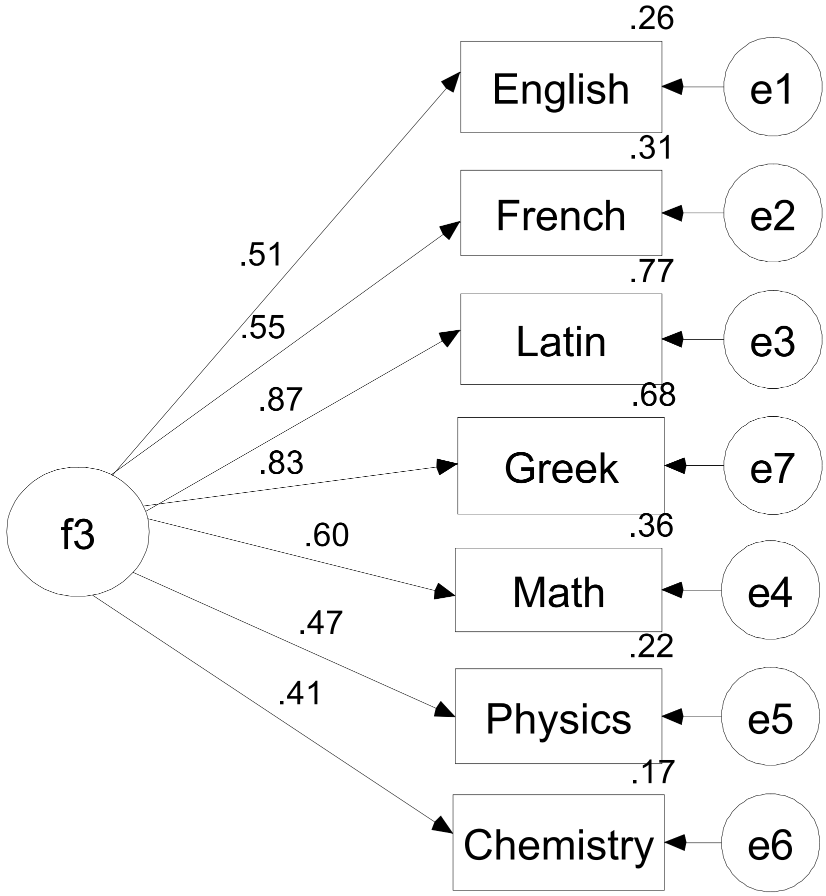
NFI=.892, CFI=.914, TLI=.835, GFI=.911

CFA 1c



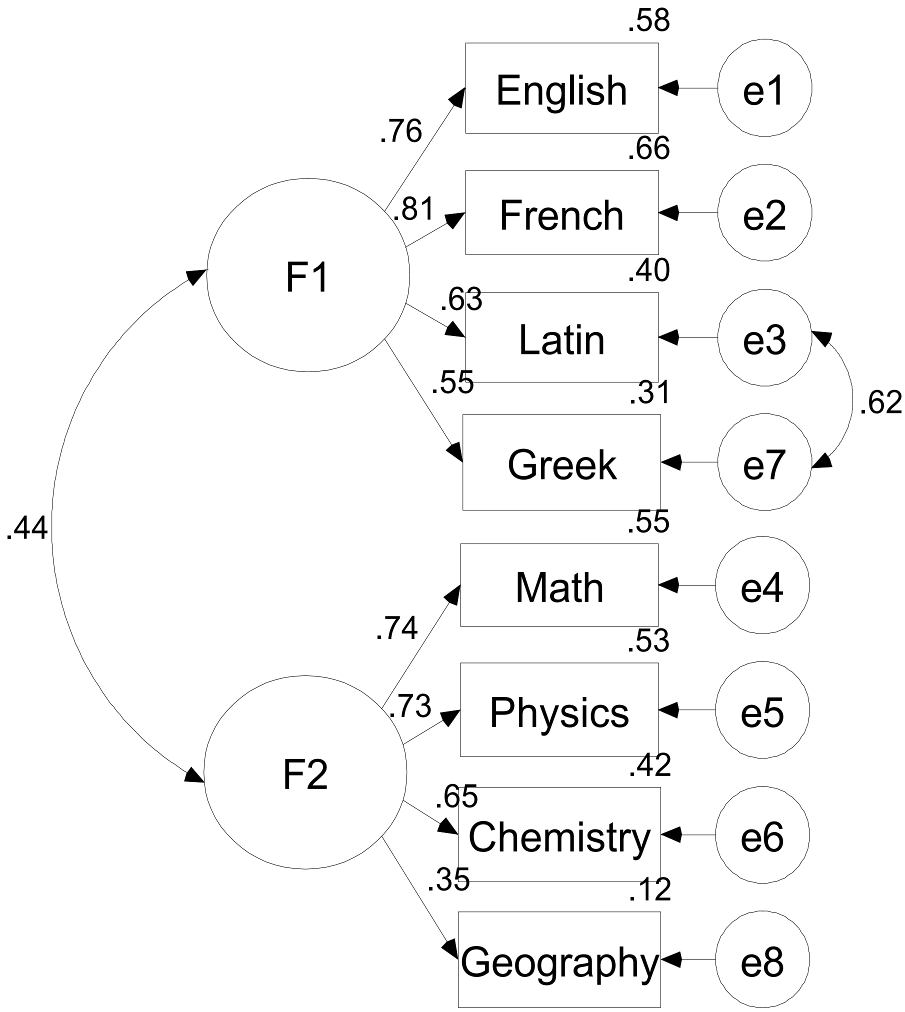
3-Factor Solution
parameters estimated: 17
ChiSquare(11, N=150)=10.684, p=.470
RMSEA=.000, RMR=.027
NFI=.975, CFI=1.000, TLI=1.002, GFI=.979

CFA 1d



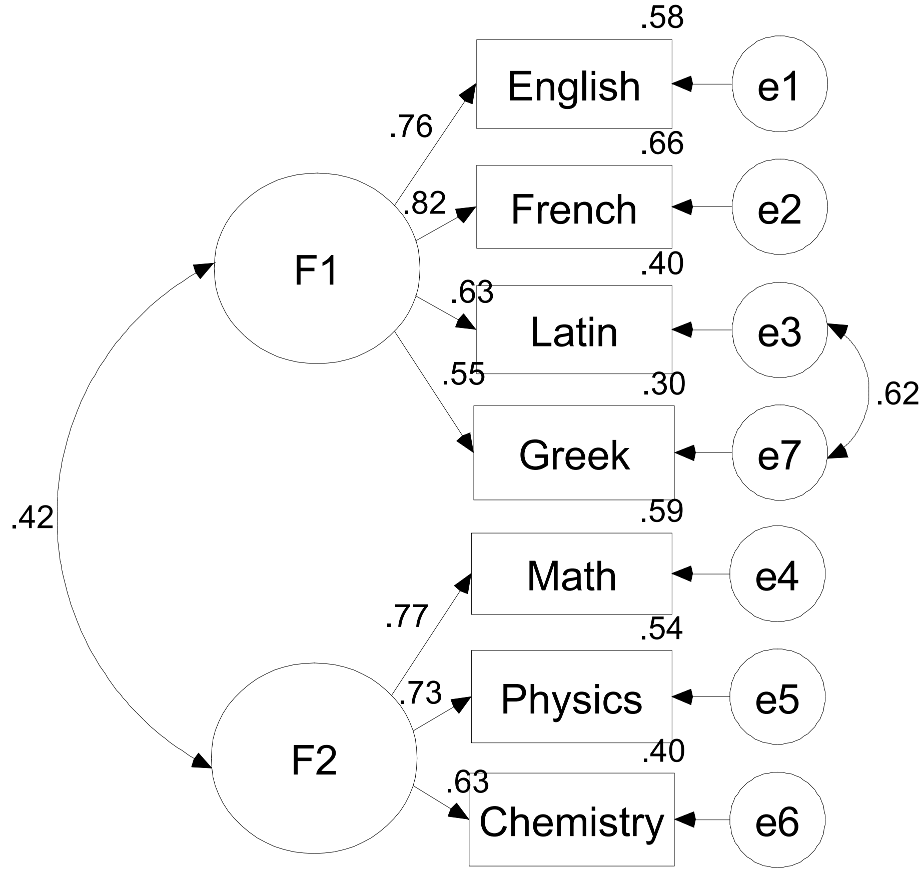
3-Factor Solution
parameters estimated: 14
ChiSquare(14, N=150)=115.668, p=.000
RMSEA=.221, RMR=.121
NFI=.724, CFI=.745, TLI=.617, GFI=.811

CFA 1e



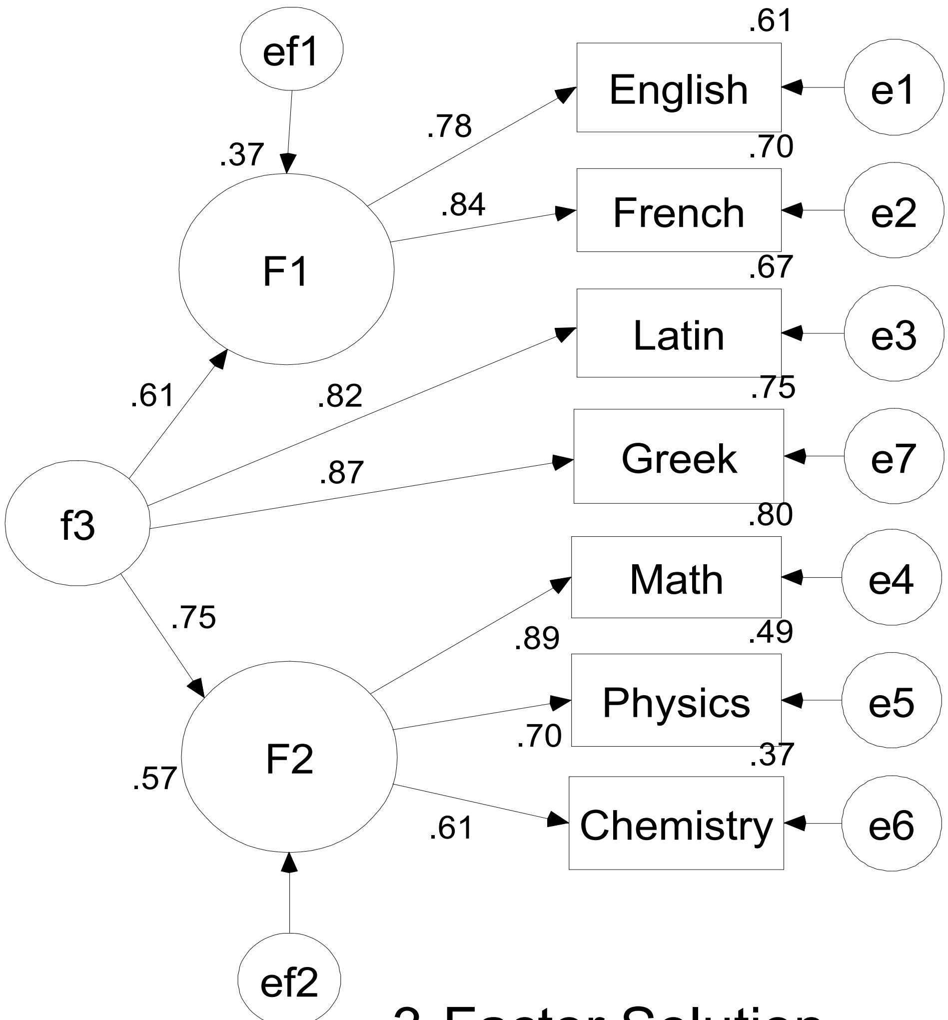
2-Factor Solution with corr errors
\# parameters estimated: 18
ChiSquare(18, N=150)=54.023, p=.000
RMSEA=.116, RMR=.104
NFI=.878, CFI=.913, TLI=.865, GFI=.920

CFA 1f



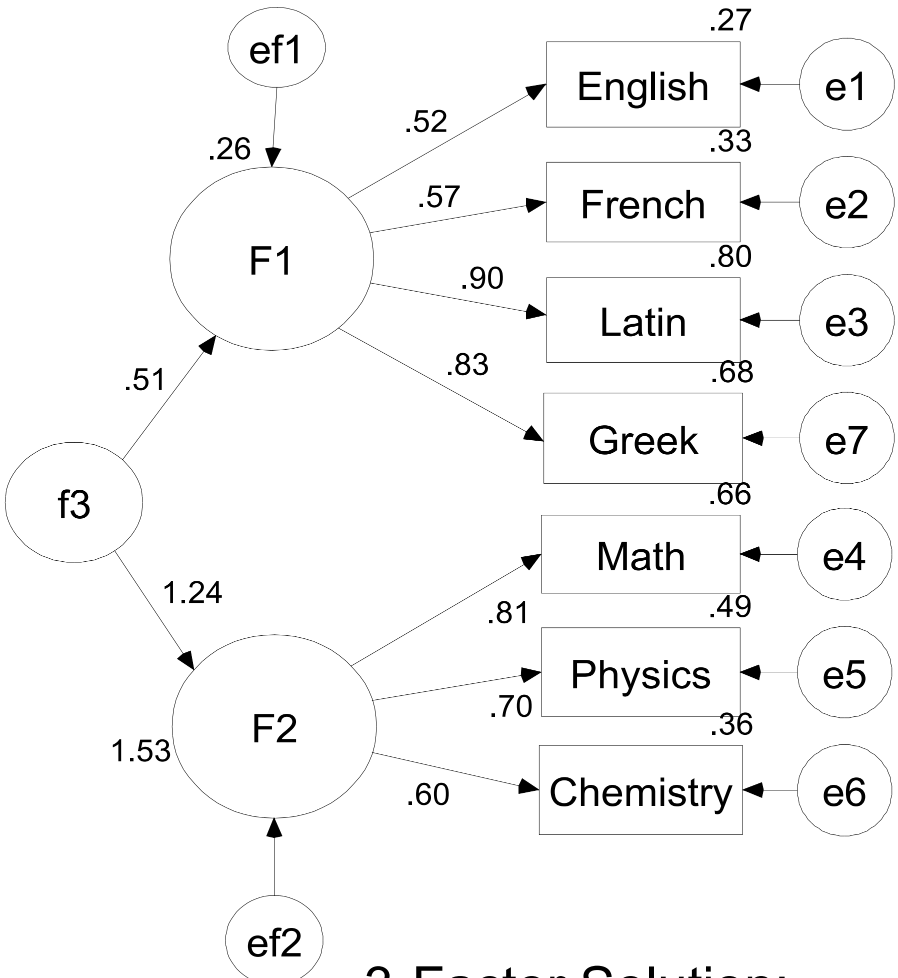
2-Factor Solution with corr error
Geography removed
parameters estimated: 16
ChiSquare(12, N=150)=43.597, p=.000
RMSEA=.133, RMR=.108
NFI=.896, CFI=.921, TLI=.861, GFI=.928

2nd order factors



3-Factor Solution
2 first order factors, 1 second order factor
parameters estimated: 15
ChiSquare(13, N=150)=29.487, p=.006
RMSEA=.092, RMR=.117
NFI=.930, CFI=.959, TLI=.933, GFI=.948

2^{nd} order factors



3-Factor Solution:
2 first order factors, 1 second order factor
parameters estimated: 15
ChiSquare(13, N=150)=60.004, p=.000
RMSEA=.156, RMR=.079
NFI=.857, CFI=.882, TLI=.809, GFI=.906