Empirical Validation of the Critical Thinking Assessment Test: A Bayesian CFA Approach

CHI HANG AU

& ALLISON AMES, PH.D.

Acknowledgement

Allison Ames, PhD

Jeanne Horst, PhD

Overview

Features of the Critical Thinking Assessment Test (CAT)

Estimation Framework
- Frequentist
- Bayesian

Methods
- Data Collection
- MCMC in Mplus

Results

Implications and Future Research

Overview

Features of the Critical Thinking Assessment Test (CAT)

Estimation Framework
- Frequentist
- Bayesian

Methods
- Data Collection
- MCMC in Mplus

Results

Implications and Future Research

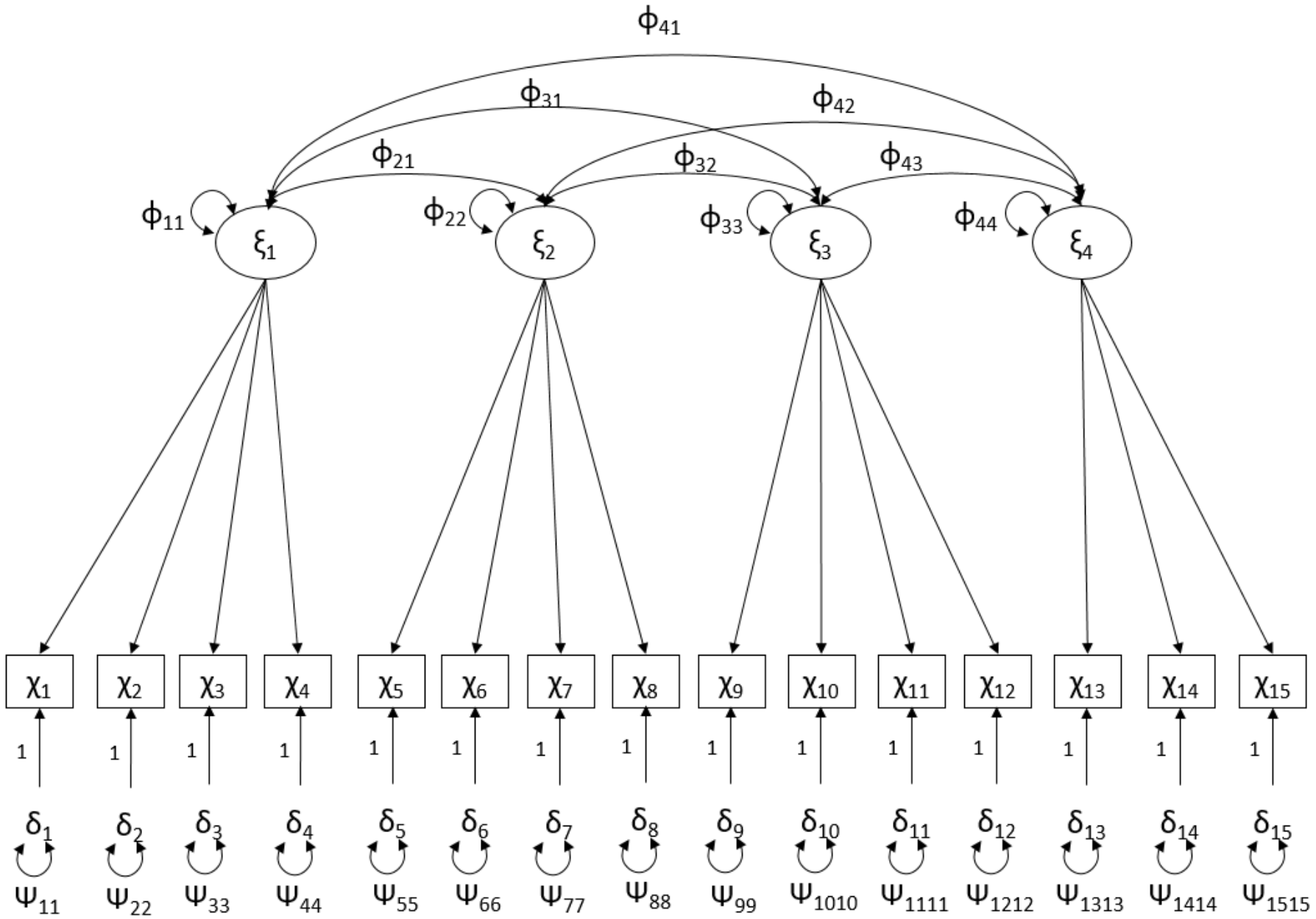
Features of the CAT

- Four domains (factors), 15 components/items (CAT; Tennessee Technological University)
  - Evaluation of Information
  - Problem Solving
  - Creative Thinking
  - Communication
- Issues:
  - Inconsistent rating scale (non-integer values)
  - Multidimensionality of items

Item	Evaluate and Interpret Information	Problem Solving	Creative Thinking	Effective Communication
Q1	**X**			
Q2	X			X
Q3			X	X
Q4		X	X	X
Q5	**X**			
Q6			X	X
Q7		X	X	X
Q8	**X**			
Q9			X	X
Q10	X	X		
Q11	X	X		X
Q12		**X**		
Q13	X	X		
Q14	X	X		X
Q15		X	X	X

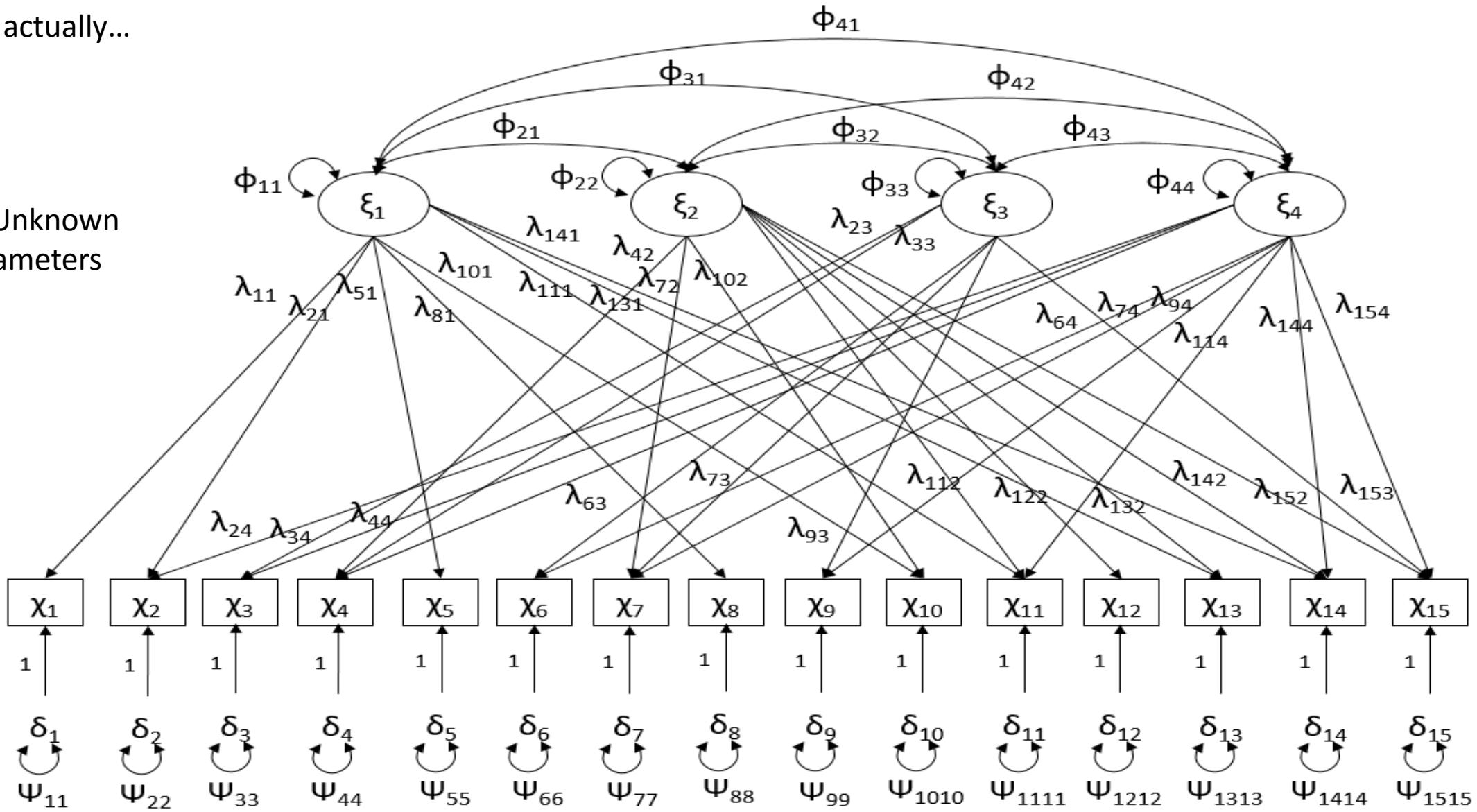
Ideal simple structure

59 Unknown Parameters



But actually…

75 Unknown Parameters



Overview

Features of the Critical Thinking Assessment Test (CAT)

Estimation Framework
- Frequentist
- Bayesian

Methods
- Data Collection
- MCMC in Mplus

Results

Implications and Future Research

Frequentist

- WLSMV
  - Relies on large sample theory (Li, 2016)
  - Local independence
  - Continuous or Categorical indicators (>5 response categories)
- Problems encountered
  - Inconsistent scoring on the CAT
  - Multidimensionality of most components

Frequentist Software Packages

- R package: lavaan – WLSMV estimation
- Mplus – WLSMV estimation

Non-convergence

lavaan

```
There were 20 warnings (use warnings() to see them)
> warnings()
Warning messages:
1: In pc_cor_TS(fit.y1 = UNI[[i]], fit.y2 = UNI[[j]], method = optim.method,  ... :
  lavaan WARNING: empty cell(s) in bivariate table of q10f x q2f
2: In pc_cor_TS(fit.y1 = UNI[[i]], fit.y2 = UNI[[j]], method = optim.method,  ... :
  lavaan WARNING: empty cell(s) in bivariate table of q14f x q2f
3: In pc_cor_TS(fit.y1 = UNI[[i]], fit.y2 = UNI[[j]], method = optim.method,  ... :
  lavaan WARNING: empty cell(s) in bivariate table of q10f x q8f
4: In pc_cor_TS(fit.y1 = UNI[[i]], fit.y2 = UNI[[j]], method = optim.method,  ... :
  lavaan WARNING: empty cell(s) in bivariate table of q11f x q10f
```

```
> fitMeasures(fit)
Error in lav_fit_measures(object = object, fit.measures = fit.measures,  :
  lavaan ERROR: fit measures not available if model did not converge
> summary(fit, fit.measures=TRUE)
** WARNING ** lavaan (0.5-22) did NOT converge after 767 iterations
** WARNING ** Estimates below are most likely unreliable
```

Mplus

Estimator	WLSMV
Maximum number of iterations	200000
Convergence criterion	0.500D-04
Maximum number of steepest descent iterations	20
Parameterization	DELTA

NO CONVERGENCE. NUMBER OF ITERATIONS EXCEEDED.

Bayesian

- Markov Chain Monte Carlo (Levy & Mislevy, 2016)
  - Use prior knowledge to guide estimation
  - Flexible, can accommodate different data type and structure
  - Does not require large sample theory or local independence

Research Questions

1. Can the CAT's factor structure be confirmed empirically using BSEM?
2. Does the BSEM approach overcome the shortcomings of other estimation methods?

Overview

Features of the Critical Thinking Assessment Test (CAT)

Estimation Framework
- Frequentist
- Bayesian

Methods
- Data Collection
- MCMC in Mplus

Results

Implications and Future Research

Data

- University-wide assessment day
- Collected from Spring 2012 to Spring 2016 (*n* = 727)
  - Missing data removed
  - Non-integer responses (disagreement among raters)
  - Final *n* = 671
- Sophomore or Junior status
  - After completing Gen Ed requirements

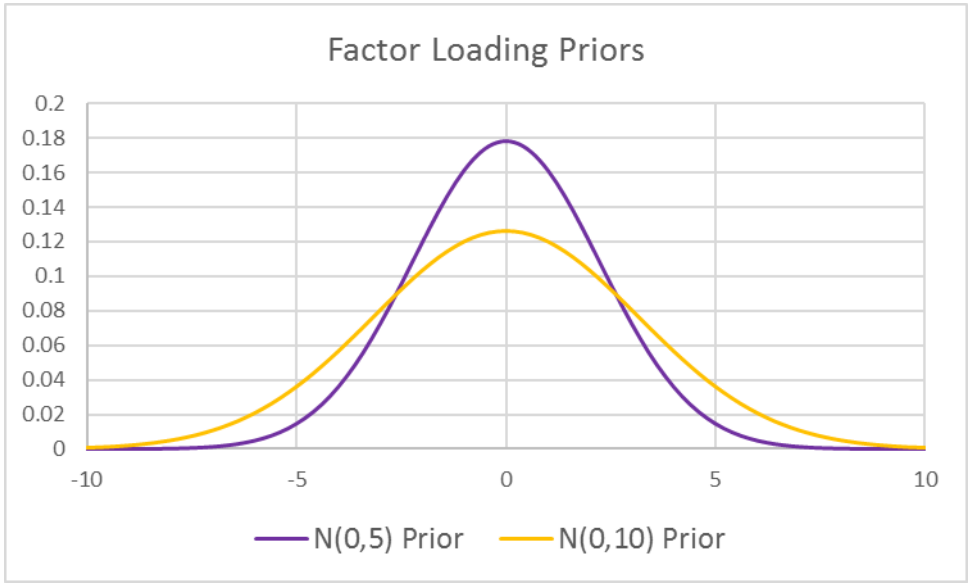
Bayesian Estimation

- Mplus Version 7.4
- Two chains
- 200,000 total iterations
  - 100,000 burn-in iterations
- ~15-20 minutes

MCMC in Mplus

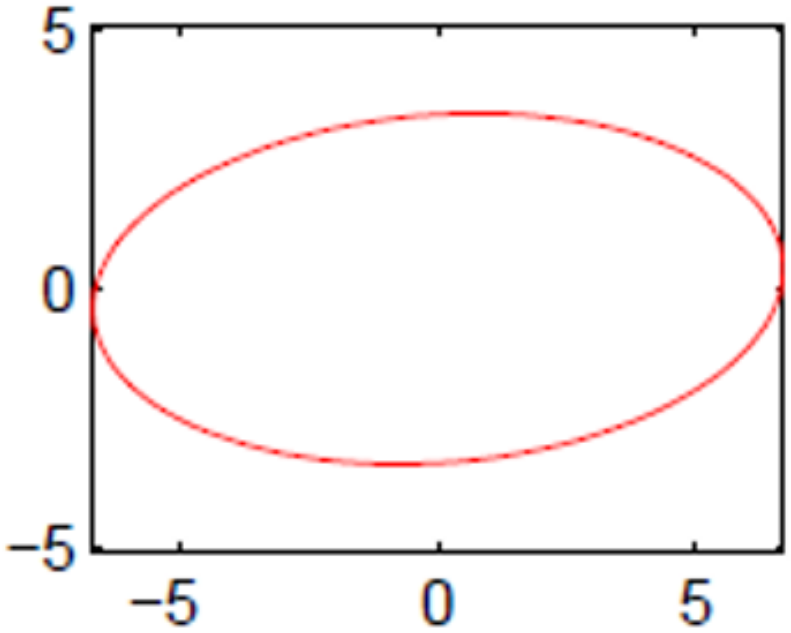
FACTOR LOADING

DEFAULT: N(0,5)



FACTOR COVARIANCE

DEFAULT: IW(0,5)



Default priors for parameter types (categorical indicators)

*Scaled Inverse-Wishart Distribution

Overview

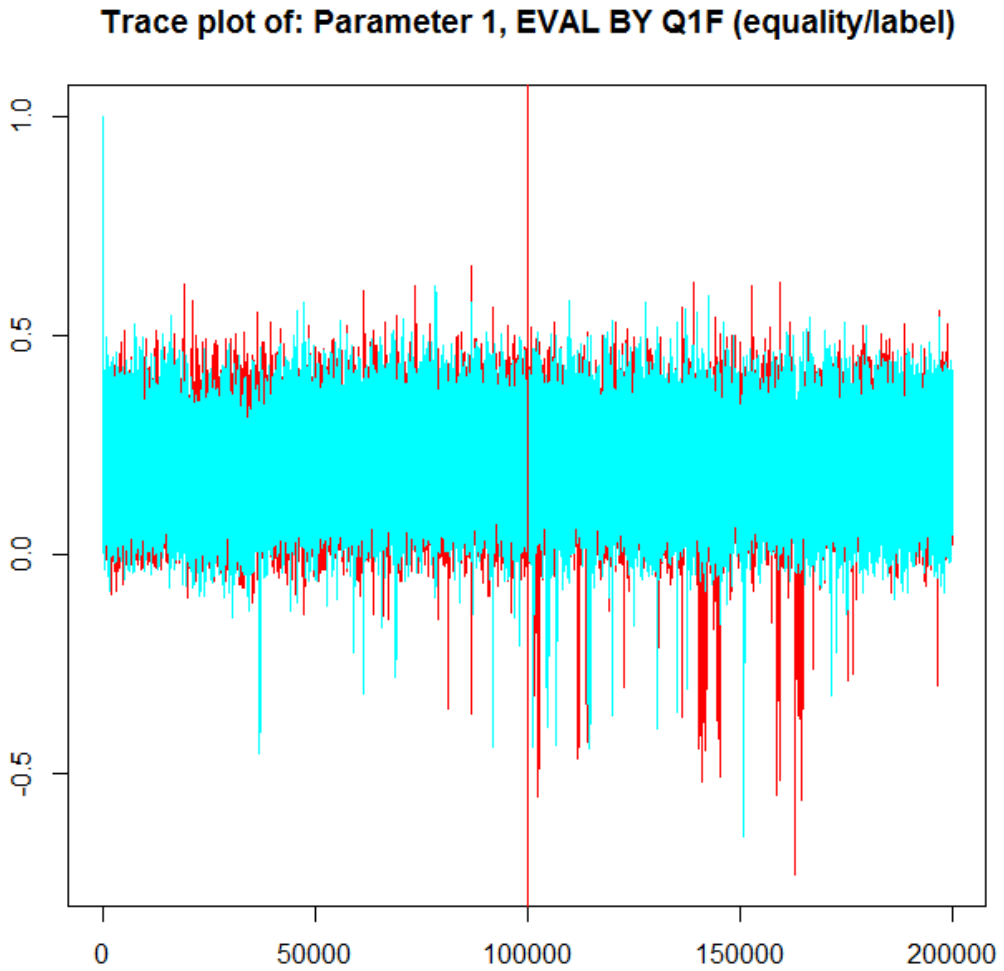
Features of the Critical Thinking Assessment Test (CAT)

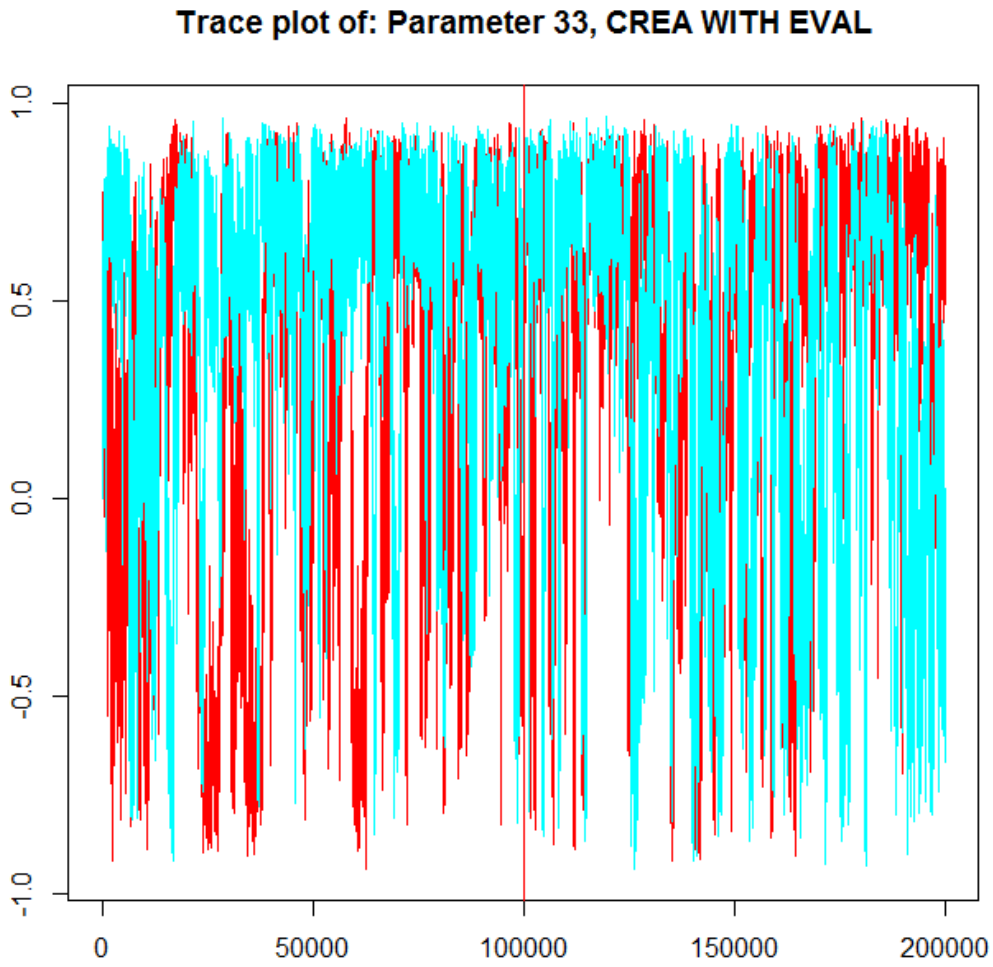
Estimation Framework
- Frequentist
- Bayesian

Methods
- Data Collection
- MCMC in Mplus

Results

Implications and Future Research


Trace plot of: Parameter 1, EVAL BY Q1F (equality/label)
1.0
0.5
0.0
-0.5
0
50000
100000
150000
200000


Trace plot of: Parameter 33, CREA WITH EVAL
1.0
0.5
0.0
-0.5
-1.0
0
50000
100000
150000
200000

Convergence

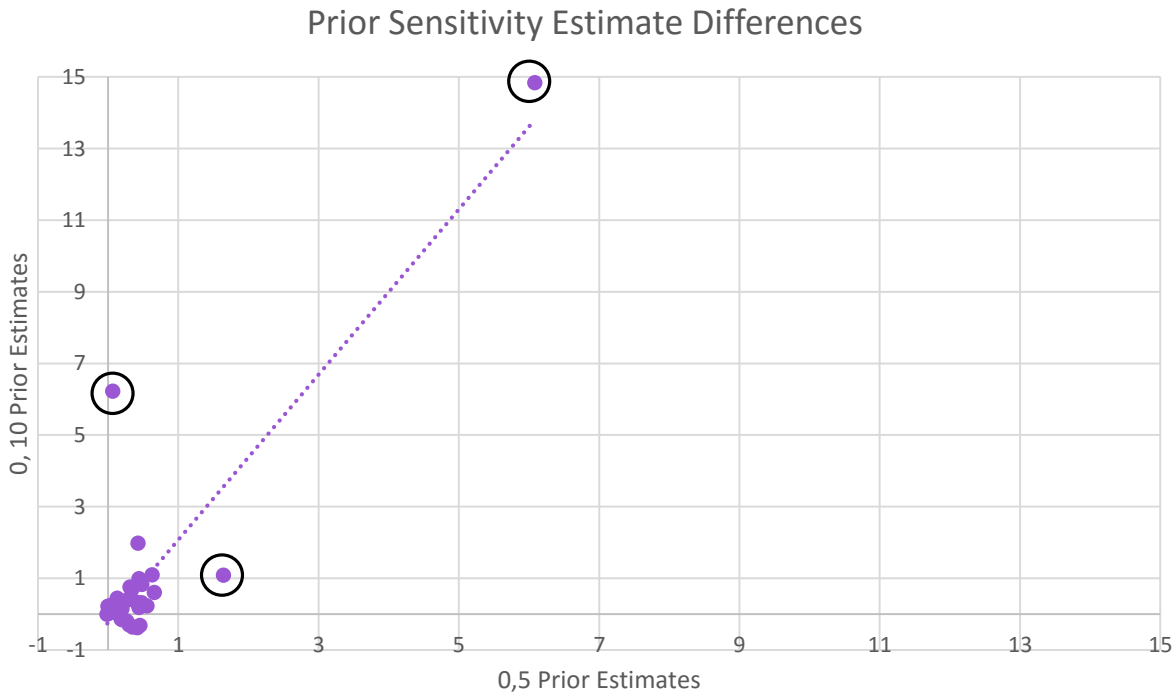
- Potential Scale Reduction (PSR) – Between chain variability by within chain variability
  - Mplus returns highest PSR value of a parameter in an iteration
  - PSR should be around 1
  - *N*(0,5): Highest PSR = 2.035 at the 200kth iteration
  - *N*(0,10): Highest PSR = 1.979 at the 200kth iteration

Model Fit – Three Methods (Gelman et al., 2003)

1. **Prior Sensitivity**: Sensitivity to different prior distributions
   - Will changing the prior substantially change the posterior?
2. **Posterior Predictive Checking**: Discrepancy measures
   - Are simulated data from the posterior data similar to the observed data?
3. **Conceptual**: Posterior inferences to substantive knowledge
   - Are the estimates or patterns consistent with theory?

Model Fit: Priors Sensitivity

- Default: *N*(0,5)
- Diffuse: *N*(0,10)
- Range of % differences
  - Q14 loads on to three factors
  - "Identify and explain the best solution for a real-world problem using relevant information"
  - Largest estimate on Problem Solving



Model Fit: Posterior Predictive Checks

- Discrepancy value for Mplus: Chi-square Statistics
- Posterior Predictive *p* – value (PPP)
  - Acceptable range: .05< *PPP* < .95
  - Obtained PPP
    - *N*(0,5) ~ .213
    - *N*(0,10) ~ .062

Model Fit: Conceptual

- Close to zero and negative interfactor correlations

Estimated Latent Factor Correlation Matrix

	1. Evaluation	2. Problem Solving	3. Creative thinking	4. Communication
1. Evaluation	1			
2. Problem Solving	-0.226	1		
3. Creative Thinking	0.593	0.052	1	
4. Communication	0.291	-0.036	-0.125	1

	Problem Solving	Creative Thinking	Communication
Q1			
Q2			X
Q3		X	X
Q4	X	X	X
Q5			
Q6		X	X
Q7	X	X	X
Q8			
Q9		X	X
Q10	X		
Q11	X		X
Q12	X		
Q13	X		
Q14	X		X
Q15	X	X	X

Model Fit: Conceptual

- Close to zero and negative interfactor correlations

Estimated Latent Factor Correlation Matrix

	1. Evaluation	2. Problem Solving	3. Creative thinking	4. Communication
1. Evaluation	1			
2. Problem Solving	-0.226	1		
3. Creative Thinking	0.593	0.052	1	
4. Communication	0.291	-0.036	-0.125	1

	Problem Solving	Creative Thinking	Communication
Q1			
Q2			X
Q3		X	X
Q4	X	X	X
Q5			
Q6		X	X
Q7	X	X	X
Q8			
Q9		X	X
Q10	X		
Q11	X		X
Q12	X		
Q13	X		
Q14	X		X
Q15	X	X	X

Model Fit: Conceptual

- Close to zero and negative interfactor correlations

Estimated Latent Factor Correlation Matrix

	1. Evaluation	2. Problem Solving	3. Creative thinking	4. Communication
1. Evaluation	1			
2. Problem Solving	-0.226	1		
3. Creative Thinking	0.593	0.052	1	
4. Communication	0.291	-0.036	-0.125	1

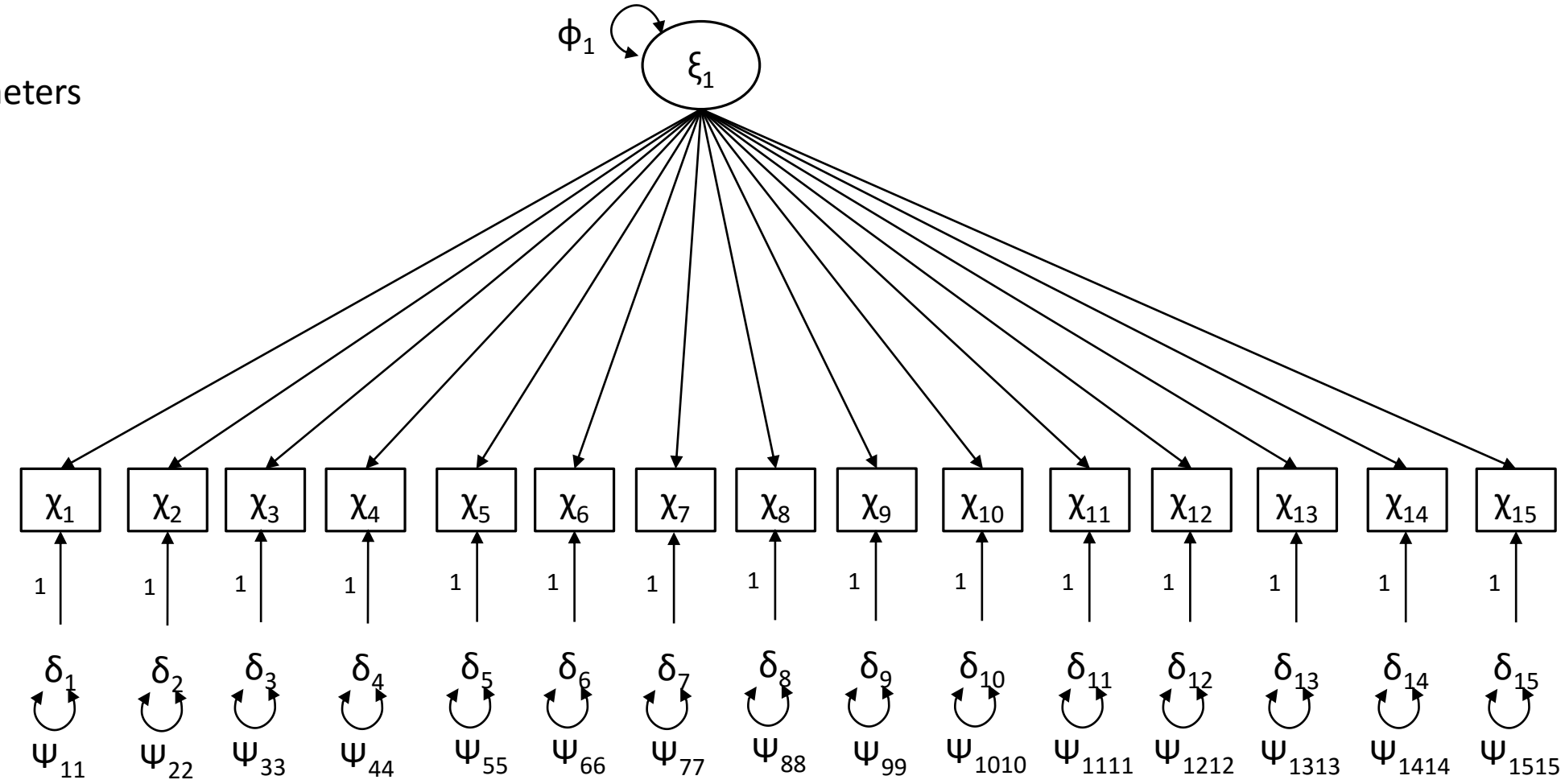
	Problem Solving	Creative Thinking	Communication
Q1			
Q2			X
Q3		X	X
Q4	X	X	X
Q5			
Q6		X	X
Q7	X	X	X
Q8			
Q9		X	X
Q10	X		
Q11	X		X
Q12	X		
Q13	X		
Q14	X		X
Q15	X	X	X

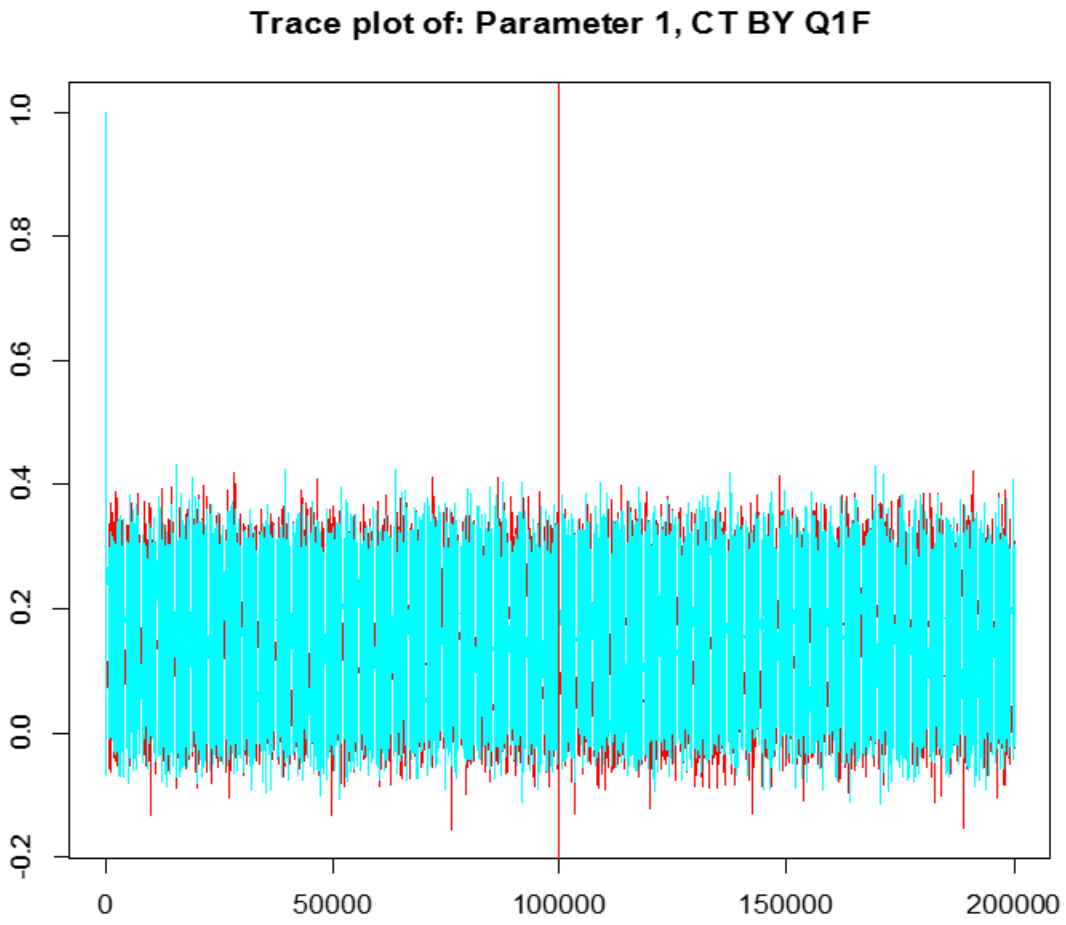
4-Factor Structure

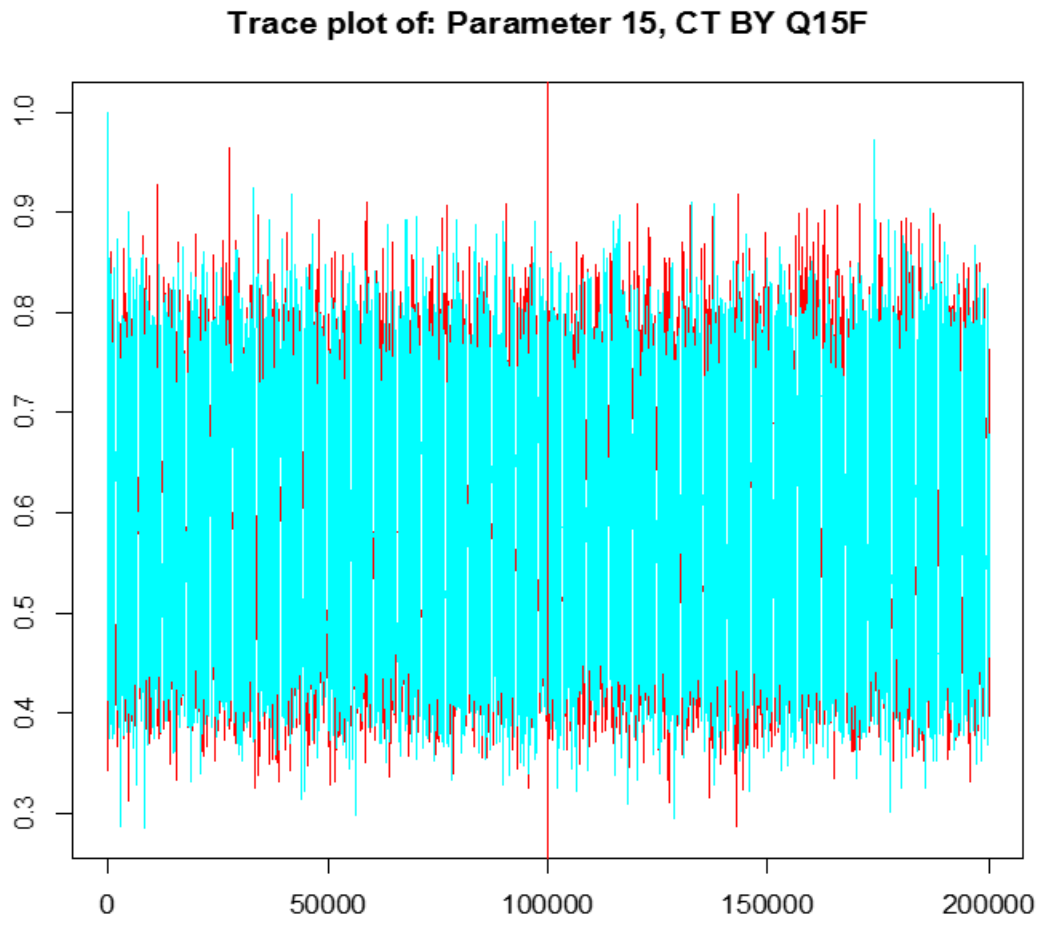
- Prior Sensitivity: Sensitivity to different prior distributions
  - % difference in parameters: -201.42% to 22,000%
  - Indicates poor model-data fit
- Posterior predictive checking
  - PPP value: ~.2
  - Indicates good model-data fit
- Conceptual: Posterior inferences to substantive knowledge
  - Close to zero and negative interfactor correlations
  - Indicates poor model-data fit

Alternative Model

1-Factor Structure
53 Unknown Parameters




Trace plot of: Parameter 1, CT BY Q1F
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
50000
100000
150000
200000

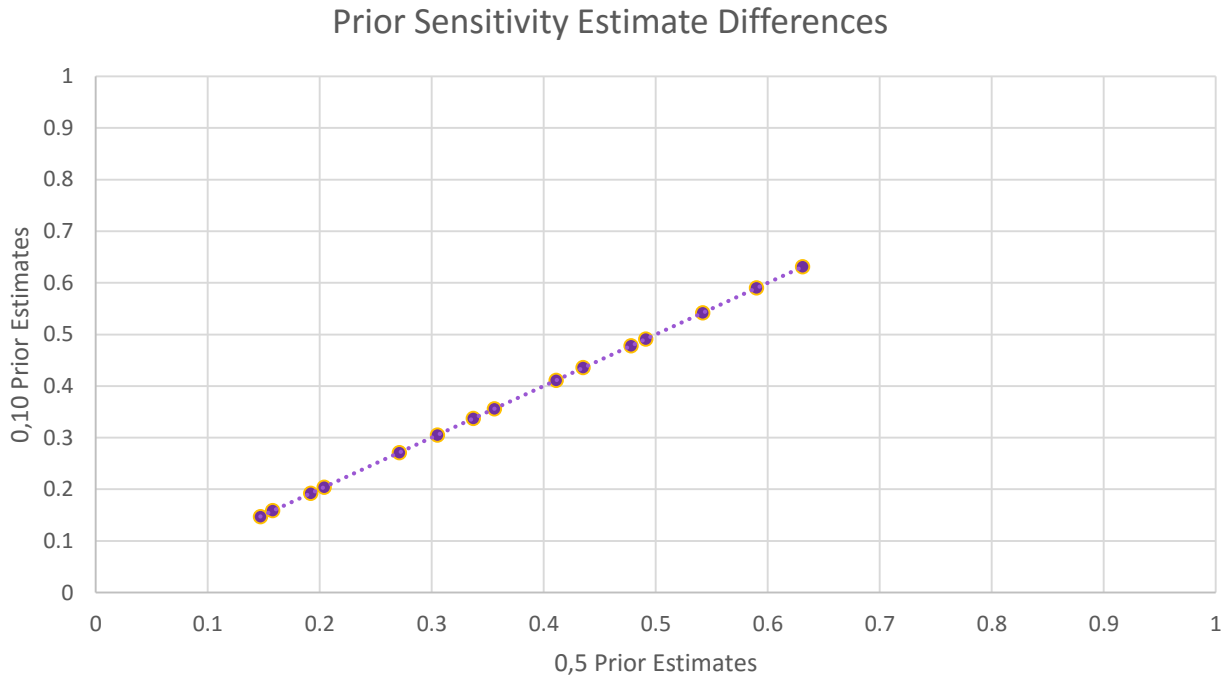

Trace plot of: Parameter 15, CT BY Q15F
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
50000
100000
150000
200000

Convergence: 1-Factor Structure

- 2 chains, 200,000 iterations
  - ~12 minutes
- *N*(0,5): Highest PSR = 1.001 at the 200kth iteration
- *N*(0,10): Highest PSR = 1.001 at the 200kth iteration

Model Fit: Priors Sensitivity

- Default: *N*(0,5)
- Diffuse: *N*(0,10)
- Range of % differences < 1%



Model Fit: Posterior Predictive Checks

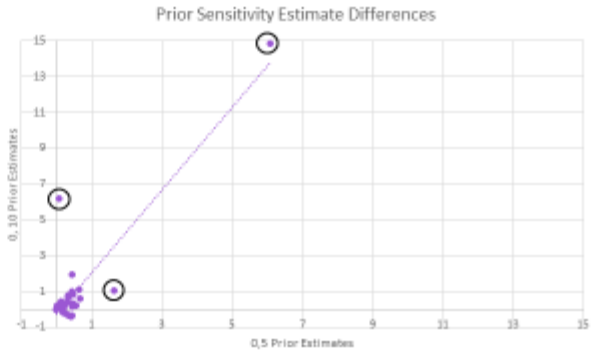
- Posterior Predictive *p* – value (PPP)
  - Acceptable range: .05< *PPP* < .95
  - Obtained PPP
    - *N*(0,5) < .001
    - *N*(0,10) < .001

Model Fit: Conceptual

- All factor loadings are non-zero and positive

Model Fit: Priors Sensitivity

- Default: *N*(0,5)
- Diffuse: *N*(0,10)
- Range of % differences
  - Q14 loads on to three factors
  - "Identify and explain the best solution for a real-world problem using relevant information"
  - Largest estimate on Problem Solving



	Critical Thinking
Q1	0.147
Q2	0.435
Q3	0.631
Q4	0.478
Q5	0.337
Q6	0.542
Q7	0.356
Q8	0.305
Q9	0.271
Q10	0.192
Q11	0.204
Q12	0.158
Q13	0.491
Q14	0.411
Q15	0.59

1-Factor Structure

- Prior Sensitivity: Sensitivity to different prior distributions
  - % difference in parameters: 0% to 0.633%
  - Indicates good model-data fit
- Posterior predictive checking
  - PPP value: <.001
  - Indicates poor model-data fit
- Conceptual: Posterior inferences to substantive knowledge
  - All factor loadings are positive (.147 to .631)
  - Indicates decent model-data fit

Overview

Features of the Critical Thinking Assessment Test (CAT)

Estimation Framework
- Frequentist
- Bayesian

Methods
- Data Collection
- MCMC in Mplus

Results

Implications and Future Research

Coming back to the Research Questions…

1. Can the CAT's factor structure be confirmed empirically using BSEM?
2. Does the BSEM approach overcome the shortcomings of other estimation methods?

Implications and Future Research

- Inconsistent evidence
- Application of BSEM approach for instrument development
- Encourage growing body of validity evidence
- Compare different factor models for the CAT
- Informative priors from content experts
- Use different discrepancy measures to assess model fit

Thank you.
Questions?

aucb@dukes.jmu.edu

ames2aj@jmu.edu

Selected References

Au, C. B. & Ames, A. J. (2016, October). *A Multitrait–Multimethod Analysis of the Construct Validity of Ethical Reasoning.* Paper presented at the annual conference of the Northeastern Educational Research Association, Trumbull, CT.

Center for Assessment and Improvement of Learning, Tennessee Technological University. https://www.tntech.edu/cat/about/

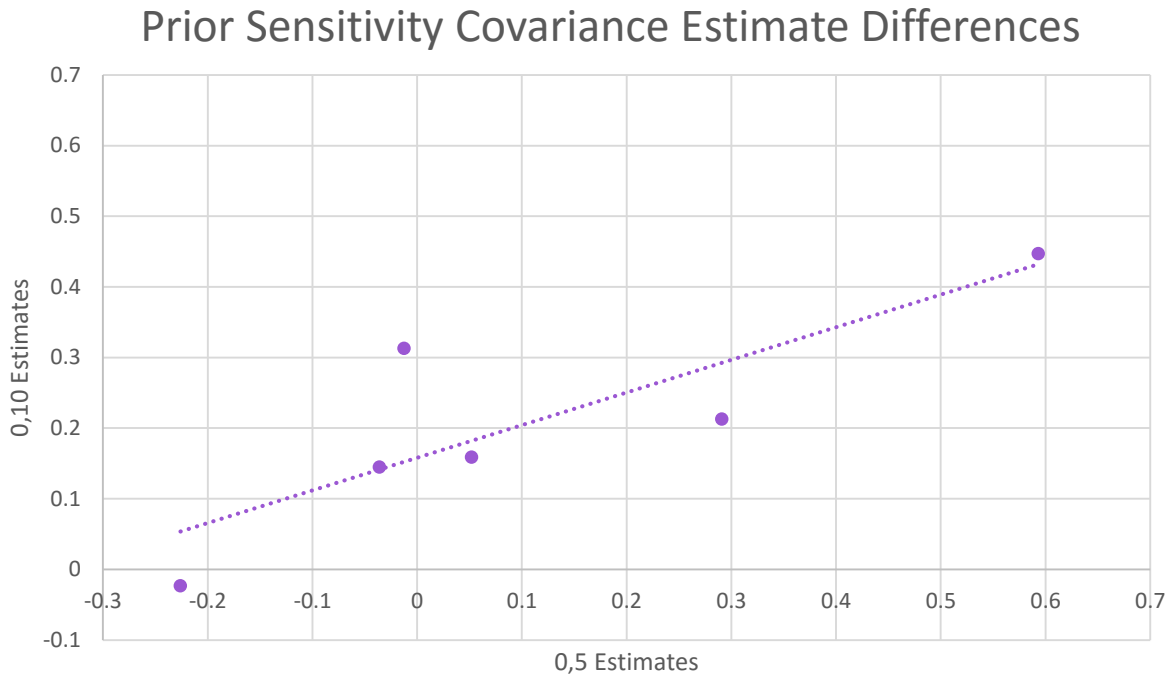
Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., Rubin, D. B. (2004). *Bayesian data analysis*. Boca Raton : CRC Press, 2004.

Levy, R. & Mislevy, R. J. (2016). Confirmatory Factor Analysis. In (Eds.)., *Bayesian Psychometric Modeling* (pp.187-230). Chapman and Hall/CRC.Center for Assessment and Improvement of Learning, Tennessee Technological University. https://www.tntech.edu/cat/about/

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods, 17*(3), 313-335.

Model Fit: Priors Sensitivity (Covariance)

- Default: *N*(0,5)
- Diffuse: *N*(0,10)
- Range of % differences
- -500% to 205%



1-Factor Results

Fit Indices			
χ^2	*p*	RMSEA	CFI
262.654	<.001	.054	.826

	Factor pattern Coefficients	Standard Error	Significance
Q1	.143	.059	.015
Q2	.386	.049	<.001
Q3	.518	.042	<.001
Q4	.415	.044	<.001
Q5	.318	.062	<.001
Q6	.466	.044	<.001
Q7	.323	.052	<.001
Q8	.296	.056	<.001
Q9	.260	.051	<.001
Q10	.186	.051	<.001
Q11	.198	.053	<.001
Q12	.164	.072	.023
Q13	.490	.044	<.001
Q14	.430	.043	<.001
Q15	.508	.043	<.001

Rating Method

- Rater effect cannot be examined
- Rater interpretation may influence multidimensionality of item-level scores

CAT Observed Correlation Matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-														
2	.181	-													
3	.045	.244	-												
4	.018	.145	.310	-											
5	.075	.181	.231	.094	-										
6	.012	.187	.281	.209	.318	-									
7	.111	.177	.186	.167	.182	.165	-								
8	.120	.152	.166	.061	.179	.210	.129	-							
9	.072	.092	.164	.115	.088	.112	.164	.304	-						
10	-.094	.084	.177	.073	.032	.112	.042	.013	-.032	-					
11	-.001	.044	.089	.069	.089	.090	.148	.076	.081	.088	-				
12	.083	.020	.036	.073	-.010	.089	.029	.213	.115	.037	.100	-			
13	.093	.158	.130	.153	.101	.184	.054	.050	.059	.070	.014	-.032	-		
14	.014	.072	.110	.117	-.007	.050	.062	.013	.051	.118	.018	.030	.492	-	
15	.062	.188	.244	.218	-.007	.225	.105	.053	.053	.048	.171	.146	.303	.345	-