Integrative Data Analysis: A Novel Methodological Framework for Data Harmonization and Pooling

Patrick J. Curran, Andrea M. Hussong, & Daniel J. Bauer

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

Psychology as a Cumulative Science

- Goal of science: build cumulative knowledge
  - "standing on shoulders of giants"
- Basis for dominant view of *logical empiricism*
  - science is systematic accrual of reproducible knowledge
- But is psychology truly a reproducible and cumulative science?

Paul Meehl (1978, p807)

“It is simply a sad fact that in soft psychology theories rise and decline, come and go, more as a function of baffled boredom than anything else; and the enterprise shows a disturbing absence of that <u>cumulative</u> character that is so impressive in disciplines like astronomy, molecular biology, and genetics.“

Recent Focus on *Reproducibility*

theguardian

Chris Chambers
Tuesday 10 June 2014

Physics envy: Do 'hard' sciences hold the solution to the replication crisis in psychology?

The physical sciences are decades – maybe centuries – ahead of psychology, but by listening and learning we have the chance to catch up

A 2012 study estimated that just 1 in 500 published psychology studies includes an exact replication of a previous experiment. Photograph: Simo Bogdanovic/Alamy

THE CHRONICLE OF HIGHER EDUCATION
August 6, 2014

Government

June 6, 2014

NIH Presses Journals to Focus on Reproducibility of Studies

By Paul Basken

A group of leading medical-journal editors, convened by the National Institutes of Health, this week endorsed a set of guidelines intended to tackle the widespread problem of scientific findings that cannot be replicated.

About 40 editors, representing journals that include *Science* and *Nature,* reached a "general agreement" about what they must accept as their responsibility for ensuring the reproducibility of their published findings, the NIH director, Francis S. Collins, said on Thursday.

Psychology Comes To Halt As Weary Researchers Say The Mind Cannot Possibly Study Itself

NEWS • Breaking • Mental Health • News • ISSUE 50•30 • Jul 31, 2014

Psychologists caution that it is a grave folly to believe anything objective can be learned about the human mind given that the object of observation is, by its nature, also the observer.

"All that we thought we understood was merely a mirage crafted by the very unfathomable minds we once so stubbornly insisted we could know," added Kaslow, before declaring the APA, with its 134,000 members and 54 academic divisions, forever disbanded.

The Quest for a Reproducible Science

- Traditional individual-sample analysis
  - limited to specific sampling frame, measures, period, etc.
- Literature reviews
  - obviously important, but risks “box score” problem
- Meta-analysis
  - powerful, but reliant on existing summary statistics
- Parallel or "*reproducibility*" analysis
  - separate analysis of raw data from independent samples
- But parallel analysis does not capitalize on *joint* characteristics of data
  - critical to examine study-specific *<u>moderators</u>* of effects

Integrative Data Analysis

- Integrative Data Analysis (IDA)
  - *the simultaneous analysis of raw data pooled from two or more independent samples* (Curran & Hussong, 2009)
- Cross-sectional, longitudinal, or some mix
- Simultaneous analysis can
  - strengthen external validity
  - enhance reproducibility
  - empirically evaluate novel research hypotheses

Potential Advantages of IDA

- Efficient use of existing resources
  - leverages data that have already been collected
  - informs coordination of data collection for ongoing studies
  - important consideration in development of new data collection

- Greater developmental age coverage
  - acceleration of time via cohort-sequential data
  - increased developmental validity in measurement
  - potential to disaggregate age, cohort and period effects
    - e.g., kids were 15 years old in 1980, 1990, and 2000

Potential Advantages of IDA

- Increased statistical power
  - larger sample sizes
  - greater sample heterogeneity & tests of subgroup differences
  - higher frequencies of low-base rate behavior

- Greater study integration and replication
  - built-in simultaneous study replication
  - direct test of study differences and moderators of effects

- Empirically test novel hypotheses in ways not possible in any single contributing data set

Potential Disadvantages of IDA

- IDA not always possible
  - incompatible measurement
  - extreme study differences
  - insufficient developmental overlap
- Challenging data management
  - often massive data sets with large numbers of items
- Complex statistical analysis
  - tractable, but many (many) steps and procedures
- But when possible, can be tremendously powerful

Typical Steps in IDA

1. Explicate theoretical question of interest
   - might be replication or novel research hypotheses
2. Identify contributing data sets
3. Develop pool of potential items
   - need common items for linking
   - all items need not be identical across all studies
4. Fit measurement model to test structure and invariance
5. Estimate optimal scores anchored to a common scale
6. Scores are then available for subsequent analysis

Common Items

- Do *not* need same measures in all contributing studies
- Can have items that are
  - identical across all studies
  - can be manually modified to be identical
  - unique within each study
- Need some subset of *common items* to establish commensurate scale for underlying construct
- Two types of common items
  1. *<u>Identical items</u>:* an item that is precisely the same in both stem and response
  2. *<u>Harmonized item</u>:* an item that has been manually modified to establish a common stem and response

Item Harmonization

- *altering an item stem or response within a study to make it comparable to similar items assessed in other studies for pooled analysis*

	Study 1	Study 2	Study 3	Harmonized item
	Consumption of alcohol			
Prompt	Over the past 6 months, on the average, how many days a month have you had a drink?	How often did you drink wine or beer or wine coolers in the past year?	Think of all the times in the past year when you had something to drink – how often have you had some kind of beverage containing alcohol?	Past-year frequency of alcohol use
Response scale	Days per month	0. Never 1. 1–2 times 2. 3–5 times 3. More than 5 times but less than once a month 4. 1–3 times a month 5. 1–2 times a week 6. 3–5 times a week 7. Every day	0. Twice a day or more 1. Once a day 2. Nearly every day 3. 3 to 4 times a week 4. Once or twice a week 5. 2 to 3 times a month 6. About once a month 7. 6–11 times a year 8. 1–5 times a year 9. Didn't drink this past year	0. Never 1. 1–5 times 2. 6–11 times 3. 1–3 times a month 4. 1–2 times a week 5. 3+ times a week

When Harmonization Doesn't Work

- Sometimes items simply can't be harmonized

	Positive expectancies about alcohol: relaxation			
Prompt	Drinking alcohol makes me relaxed	Drinking alcohol relaxes me	Drinking helps me to relax	Expectation that alcohol helps to relax
Response scale	0. Never 1. Very rarely 2. Rarely 3. Occasionally 4. Frequently 5. Very frequently 6. Always	1. Strongly agree 2. Agree 3. Neither agree nor disagree 4. Disagree 5. Strongly disagree	1. Not at all 2. A little bit 3. Somewhat 4. Quite a bit 5. A lot	?

Harmonization Alone is Insufficient

- Even if successful, cannot assume that either common <u>or</u> harmonized items are equivalent across person or study
- Harmonized values may:
  - understate alcohol use in free format, but not in intervals
  - introduce variation due to differing item prompts, response labels or battery placement
- May introduce <u>*artifacts*</u> into analysis that really due to study differences in stem or response
- Can use psychometric models to formally evaluate these study-specific differences

Traditional Psychometric Models

- Traditionally, measurement invariance examined by confirmatory factor analysis and item response theory
- Excellent approaches, not always ideal for IDA
  - difficulty including mixed scale types
  - invariance tests limited to discrete group membership
- Recent analytic development avoids limitations
  - moderated nonlinear linear factor analysis (MNLFA) model (Bauer & Hussong, 2009)
- Will demonstrate MNLFA to obtain scores using 17 binary items assessing depression over time

Motivating Example: Cross Study

- NIDA-funded project combines 3 existing data sets to study pathways to substance use
  - Michigan Longitudinal Study (MLS; Bob Zucker)
  - Adolescent Family Development Project (AFDP; Laurie Chassin)
  - Alcohol & Health Behavior Project (AHBP; Ken Sher)
- *Brief* exemplar goal for today
  - create individual- & time-specific scores of depression using 17 items from pooled sample where no study assessed all items
  - estimated scores can then be used in subsequent modeling

Cross Study Design

<table>
<tr><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th><th>11</th><th>12</th><th>13</th><th>14</th><th>15</th><th>16</th><th>17</th><th>18</th><th>19</th><th>20</th><th>21</th><th>22</th><th>23</th><th>24</th><th>25</th><th>26</th><th>27</th><th>28</th><th>29</th><th>30</th><th>31</th><th>32</th><th>33</th><th>34</th><th>35</th><th>36</th><th>37</th><th>38</th><th>39</th><th>40</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="4">MLS- battery 9, n=12</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="6">MLS- battery 8, n=62</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="10">MLS-battery 7, n=130</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="5">MLS- battery 6, n=57</td><td></td><td></td><td></td><td></td><td></td><td colspan="14">AFDP- battery 6, cohort 2 n=331</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="5">MLS- battery 5, n=499</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="9">AFDP- battery 6, cohort 1 n=390</td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="5">MLS-battery 4, n=510</td><td></td><td></td><td></td><td></td><td></td><td colspan="14">AFDP- battery 5, cohort 2 n=345</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="5">MLS- battery 3, n=431</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="10">AFDP- battery 5, cohort 1 n=411</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td colspan="6">MLS- battery 2, n=349</td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="11">AFDP- battery 4, cohort 2=327</td><td></td><td></td><td></td><td></td><td></td><td colspan="5">AHBP-battery 7, n=380</td><td></td><td></td><td></td></tr>
<tr><td colspan="6">MLS- battery 1, n=371</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="7">AFDP- battery 4, cohort 1=422</td><td></td><td></td><td></td><td></td><td colspan="6">AHBP- battery 6, n=406</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="7">AFDP- battery 3, cohort 1 n=447</td><td></td><td></td><td></td><td></td><td colspan="7">AHBP- battery 5, n=454</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="7">AFDP- battery 2, cohort 1 n=449</td><td></td><td></td><td></td><td></td><td colspan="4">AHBP-battery 4, n=467</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="7">AFDP- battery 1, cohort 1 n=454</td><td></td><td></td><td colspan="7">AHBP- battery 3, n=468</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="7">AHBP- battery 2, n=480</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="7">AHBP- battery 1, n=485</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Integrated Sample for Model Fitting

	MLS (n=641)	AFDP (n=846)	AHBP (n=485)	Pooled Repeated Measures Sample (n=1972)
Age	15.24(3.12)	21.17(7.05)	22.79(4.77)	19.81(6.21)
% Male	71.0	52.4	47.2	57.2
% COA	76.0	50.4	48.7	58.3
% Minority	2.3	30.3	6.2	15.3
% Parent ASP	14.8	9.6	7.8	11.0
%Parent Depression	24.6	16.8	36.3	25.0
Parent Education	2.59(1.18)	3.09(1.13)	3.62(1.14)	3.05(1.21)

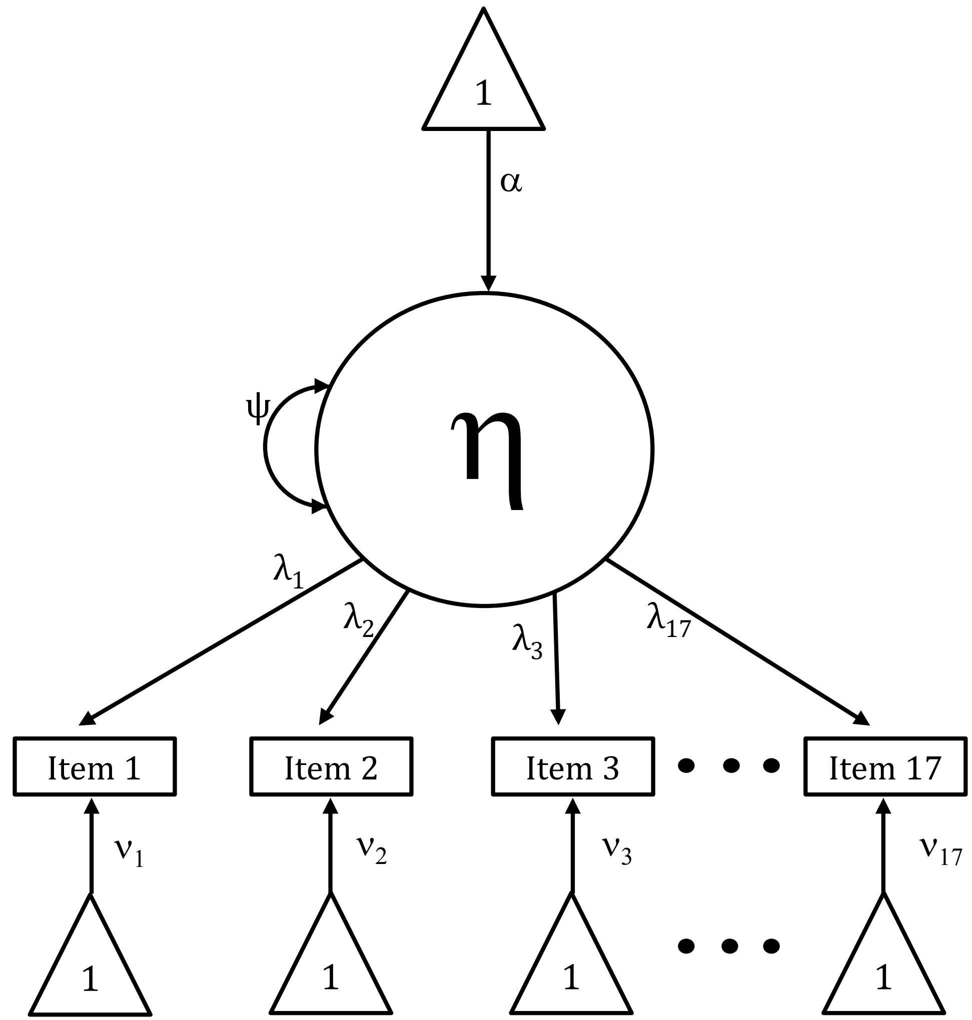
Pool of Available Items

- 33 binary self-report items assessing presence or absence of internalizing symptomatology ages 11-35
  - some from Brief Symptom Inventory (BSI)
  - some from Child Behavior Check List (CBCL)
  - some items share content across BSI and CBCL
- Variations in item coverage across studies
  - MLS: all items administered
  - AFDP: subset of CBCL items administered
  - AHBP: all BSI items administered
- Subset of common items in all studies allow for linking and unique items within-study increase score precision
- Preliminary EFAs identified 17 items defining *depression*

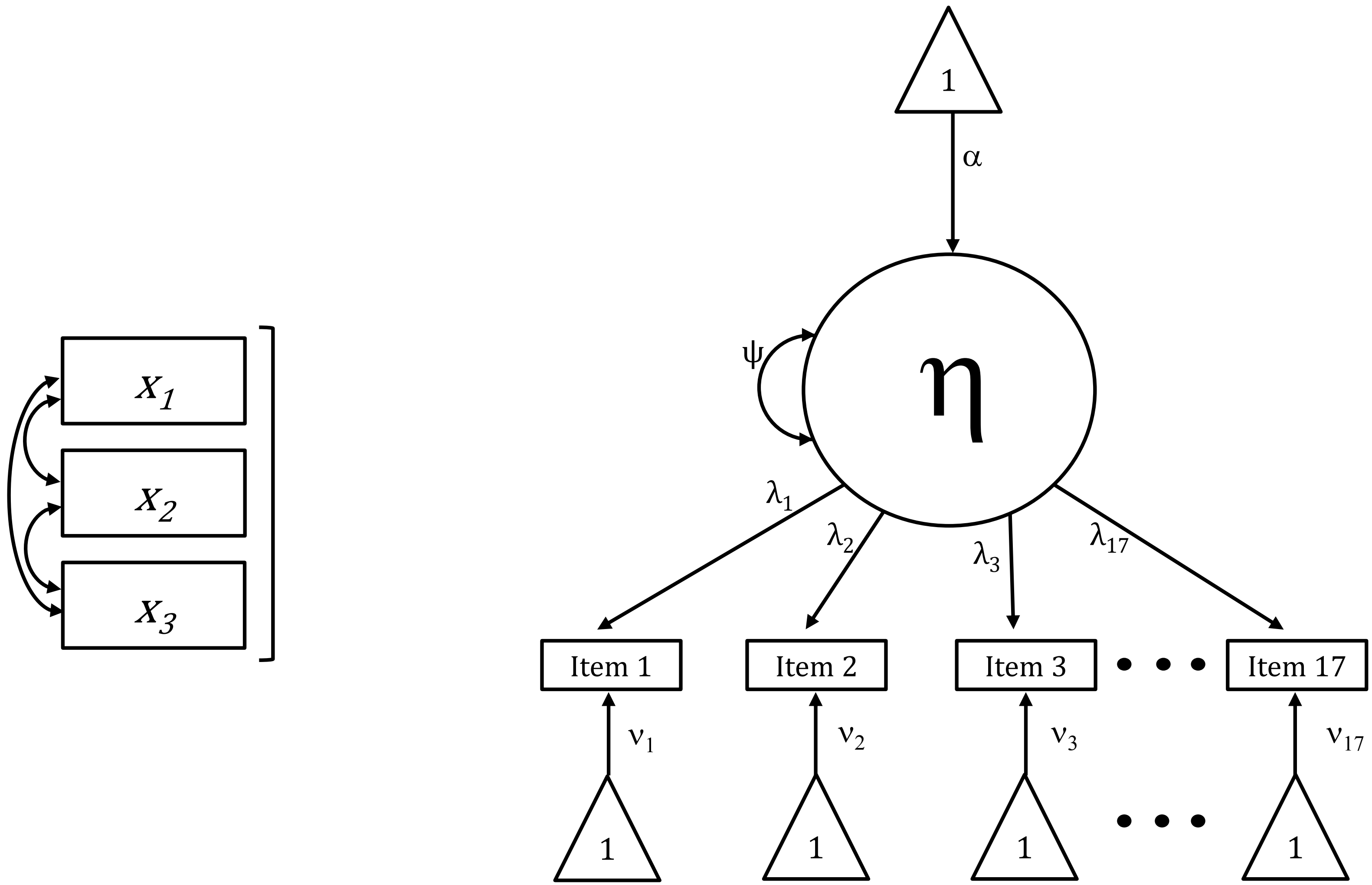
EFA Results for 17 Depression Items

Item Description	Loading (*se*)	Intercept (*se*)
1. Lonely	2.53 (.17)	-0.74 (.09)
2. Cries a lot	1.52 (.12)	-1.60 (.10)
3. Fears will behave badly	1.33 (.17)	-1.61 (.13)
4. Have to be perfect	1.16 (.09)	-0.04 (.07)
5. No one loves me	2.50 (.21)	-3.26 (.21)
6. Worthless/inferior	2.85 (.21)	-2.98 (.19)
7. Prefers being alone	0.96 (.13)	-0.52 (.10)
8. Feel guilty	1.70 (.11)	-1.70 (.09)
9. Is secretive	1.51 (.17)	0.31 (.11)
10. Is underactive	1.11 (.14)	-0.76 (.10)
11. Unhappy/ sad/depressed	2.61 (.20)	-0.67 (.10)
12. Worried	1.82 (.13)	0.23 (.08)
13. Hopeless about future	2.16 (.21)	-1.99 (.15)
14. Acts to harm self	1.75 (.36)	-4.30 (.43)
15. Thinks about killing self	1.83 (.24)	-3.76 (.27)
16. Blue	2.54 (.23)	-0.46 (.11)
17. No interest in things	1.62 (.15)	-0.99 (.10)

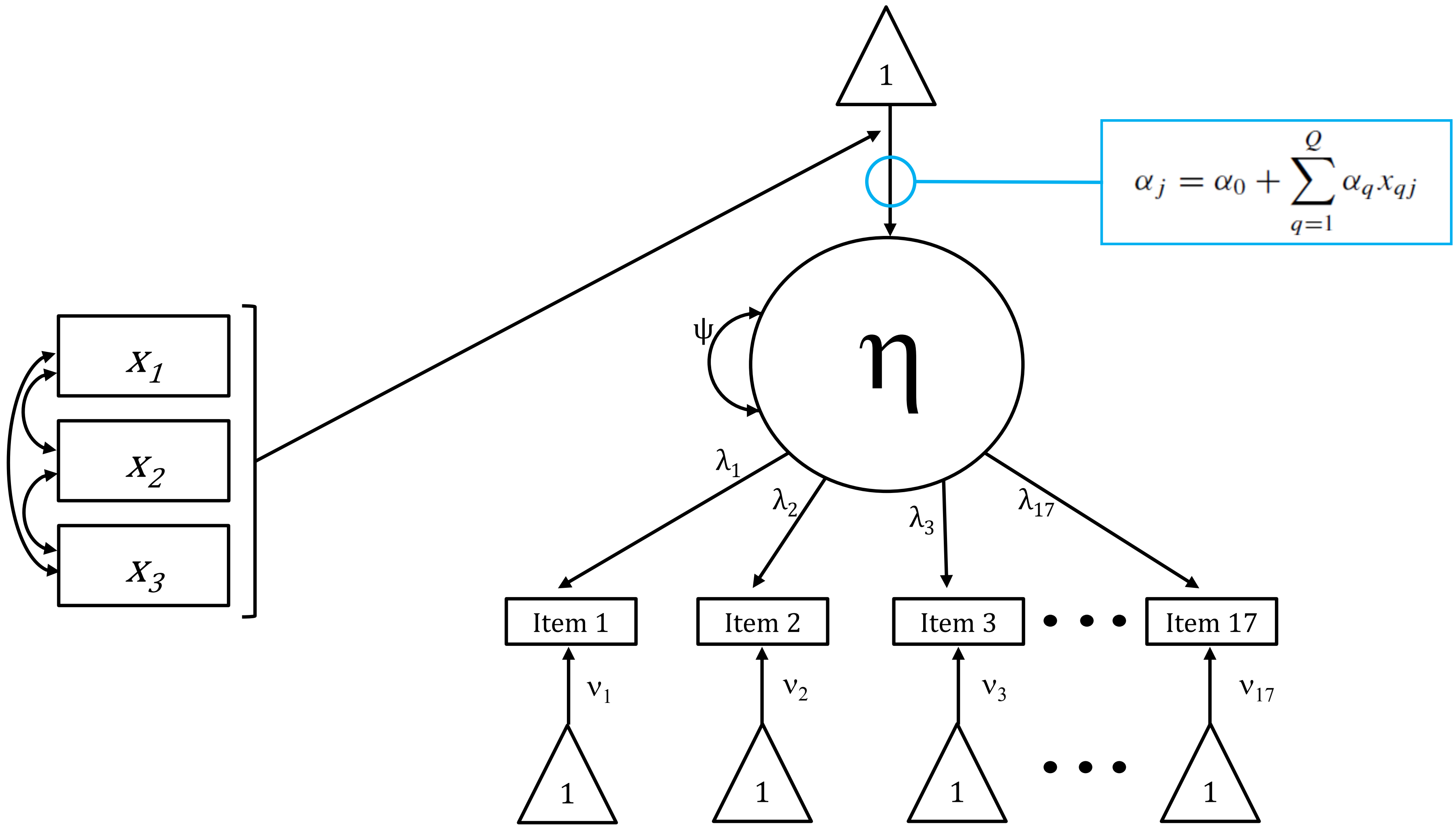
Nonlinear Confirmatory Factor Analysis



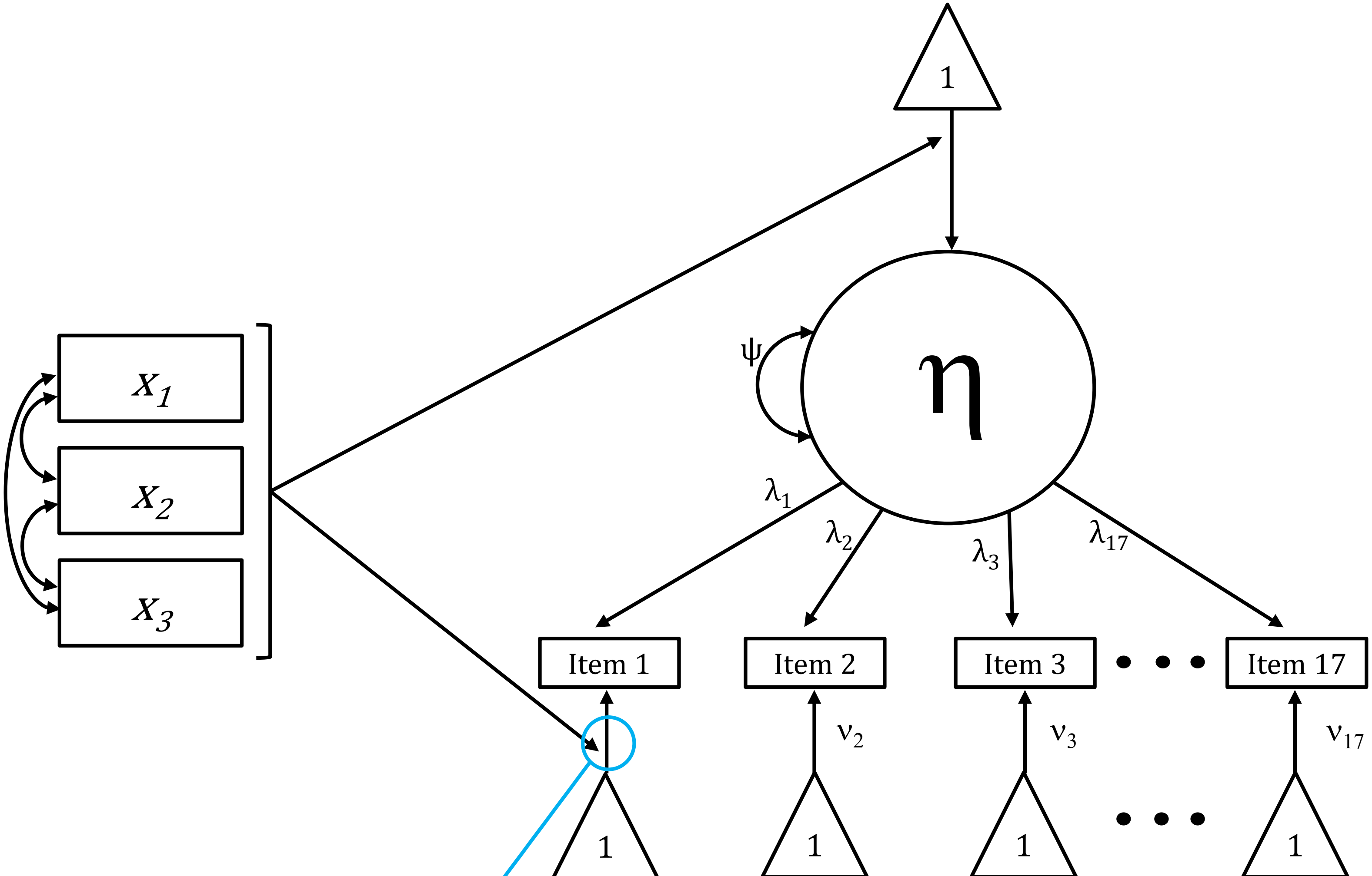
CFA with Exogenous Covariates



Factor Mean Differences

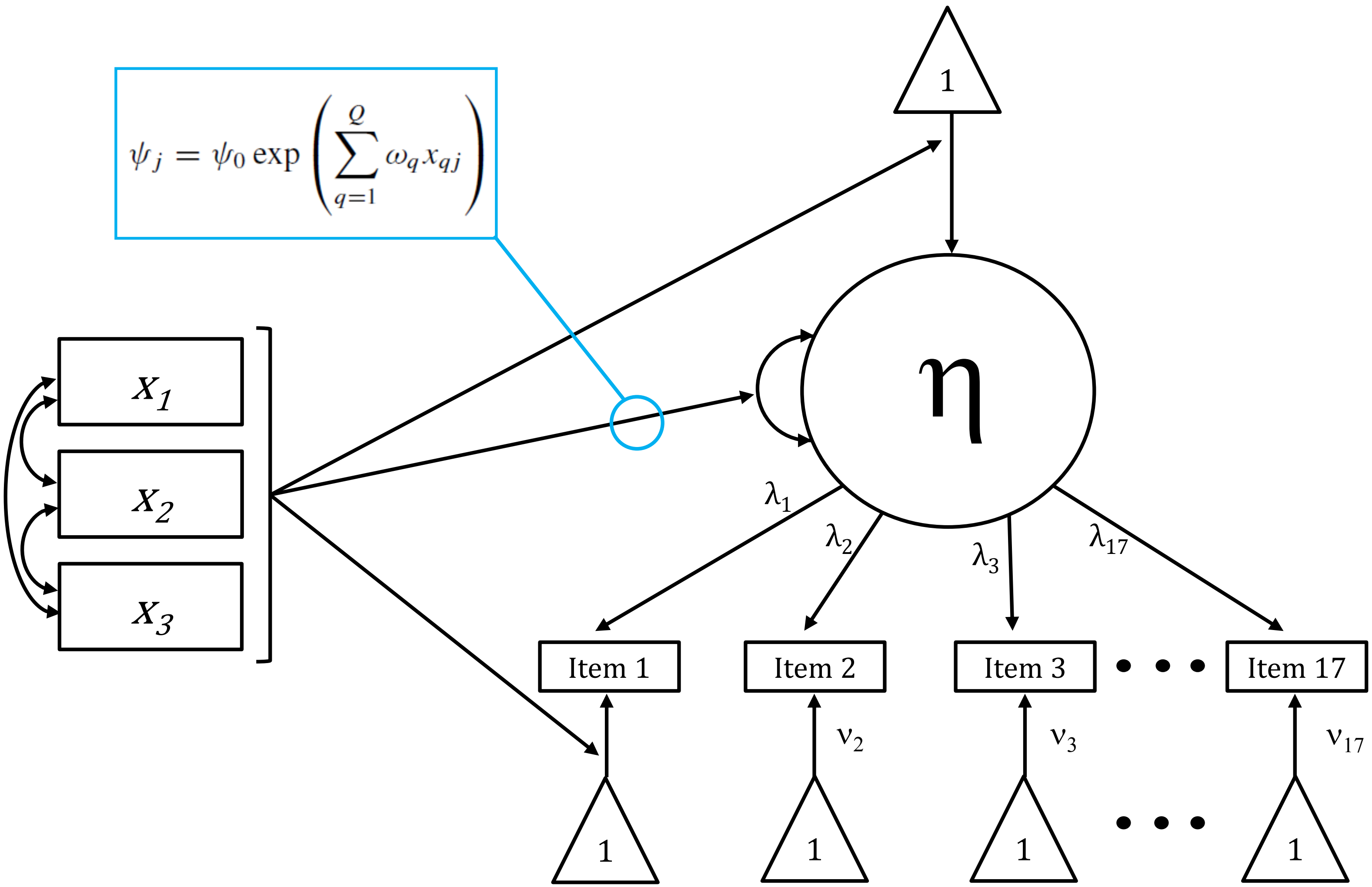


Item Intercept Differences

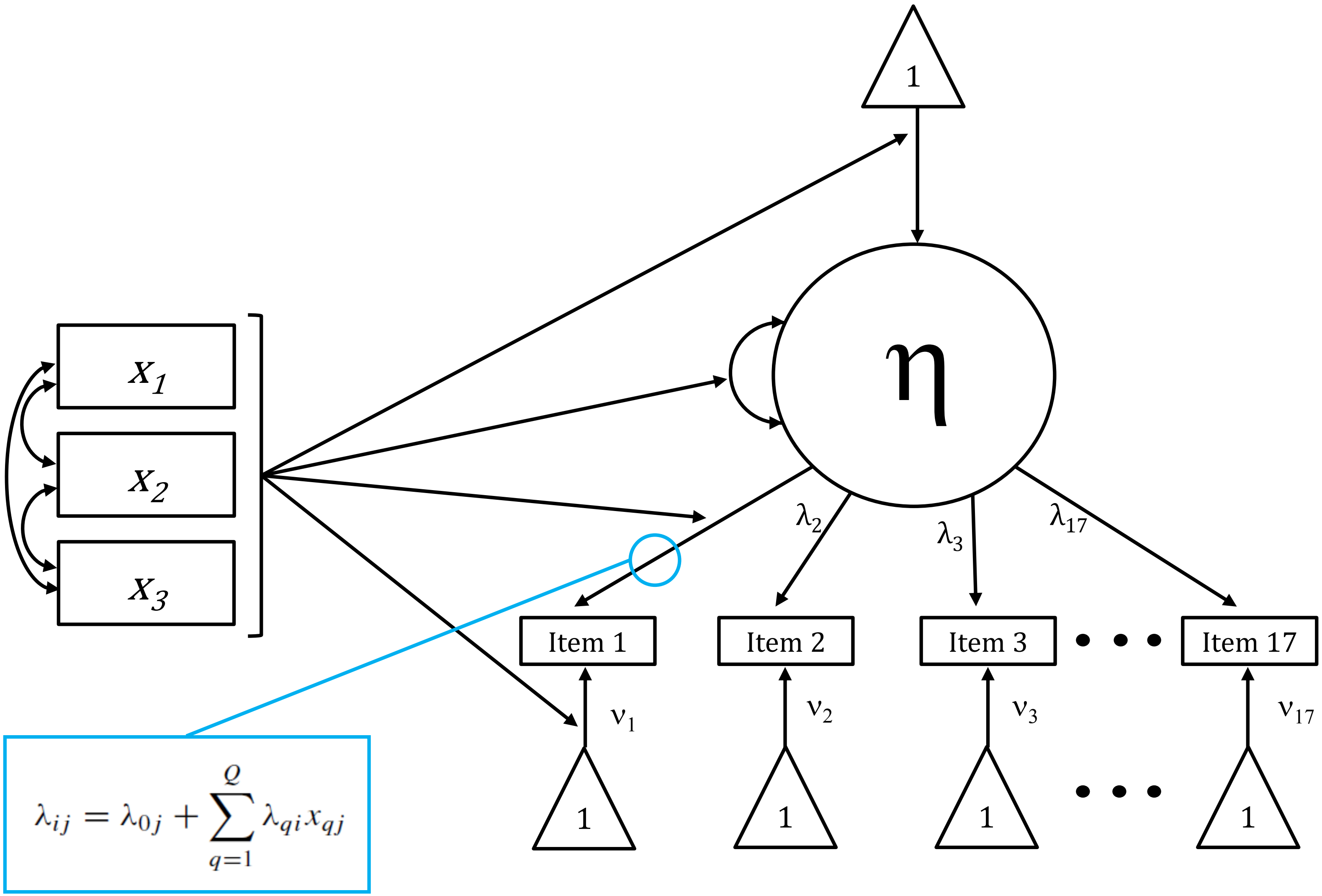


$$\nu_{ij} = \nu_{0i} + \sum_{q=1}^{Q} \nu_{qi} x_{qj}$$

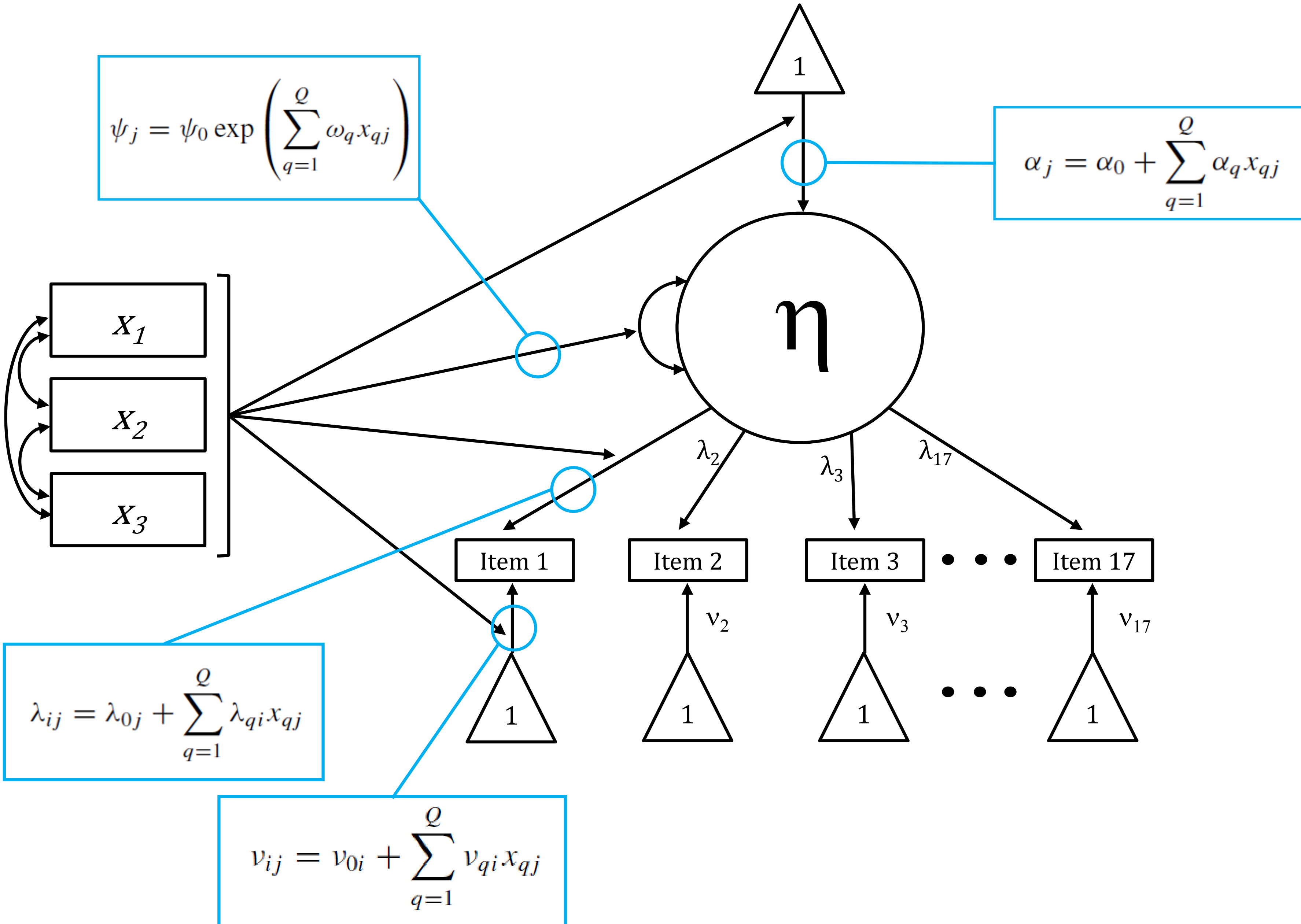
Factor Variance Differences

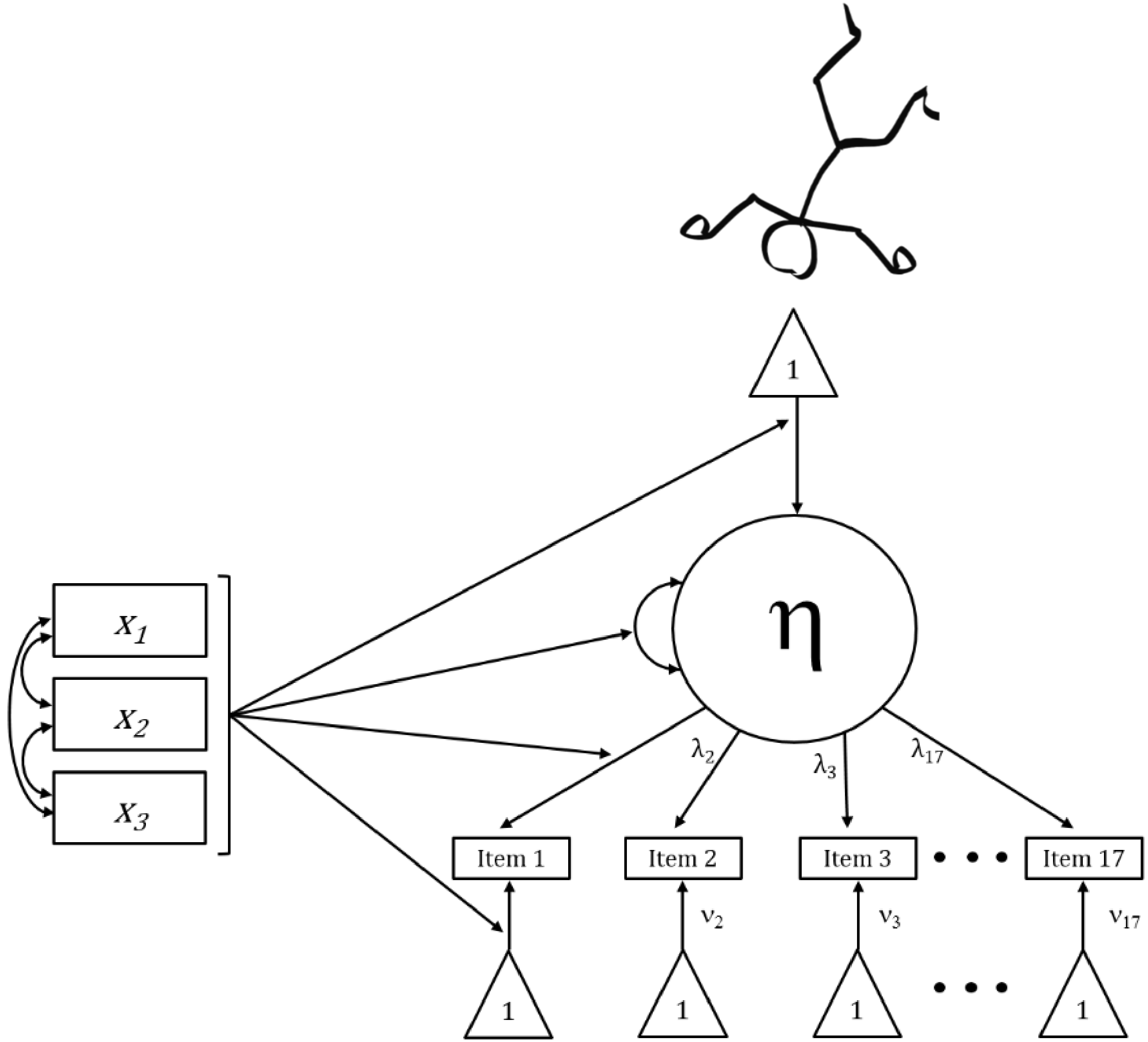


Factor Loading Differences



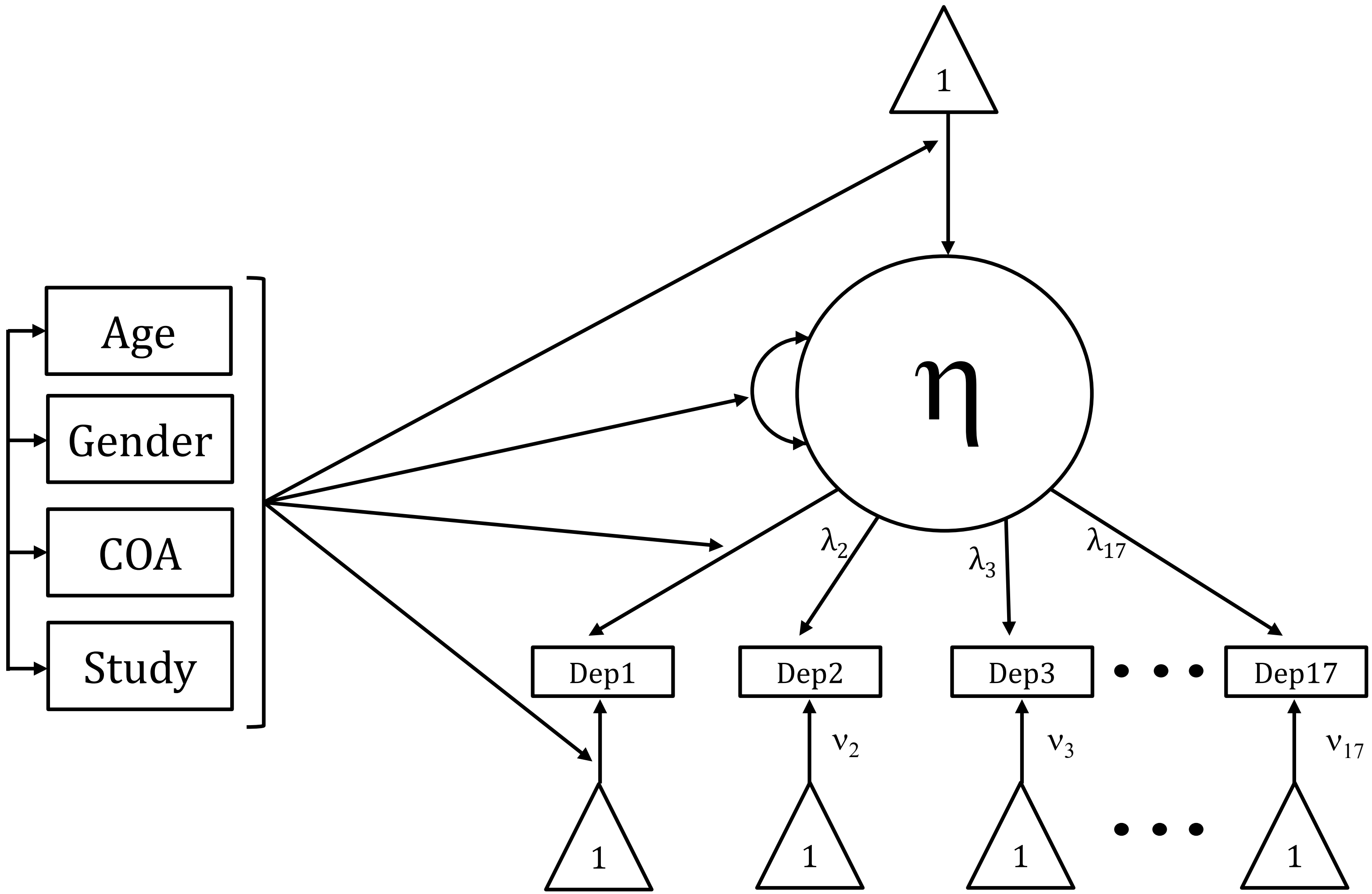
Full Moderated Nonlinear Factor Analysis





1
η
X_1
X_2
X_3
λ_2
λ_3
λ_{17}
Item 1
Item 2
Item 3
Item 17
ν_2
ν_3
ν_{17}
1
1
1
1

MNLFA with Exogenous Covariates



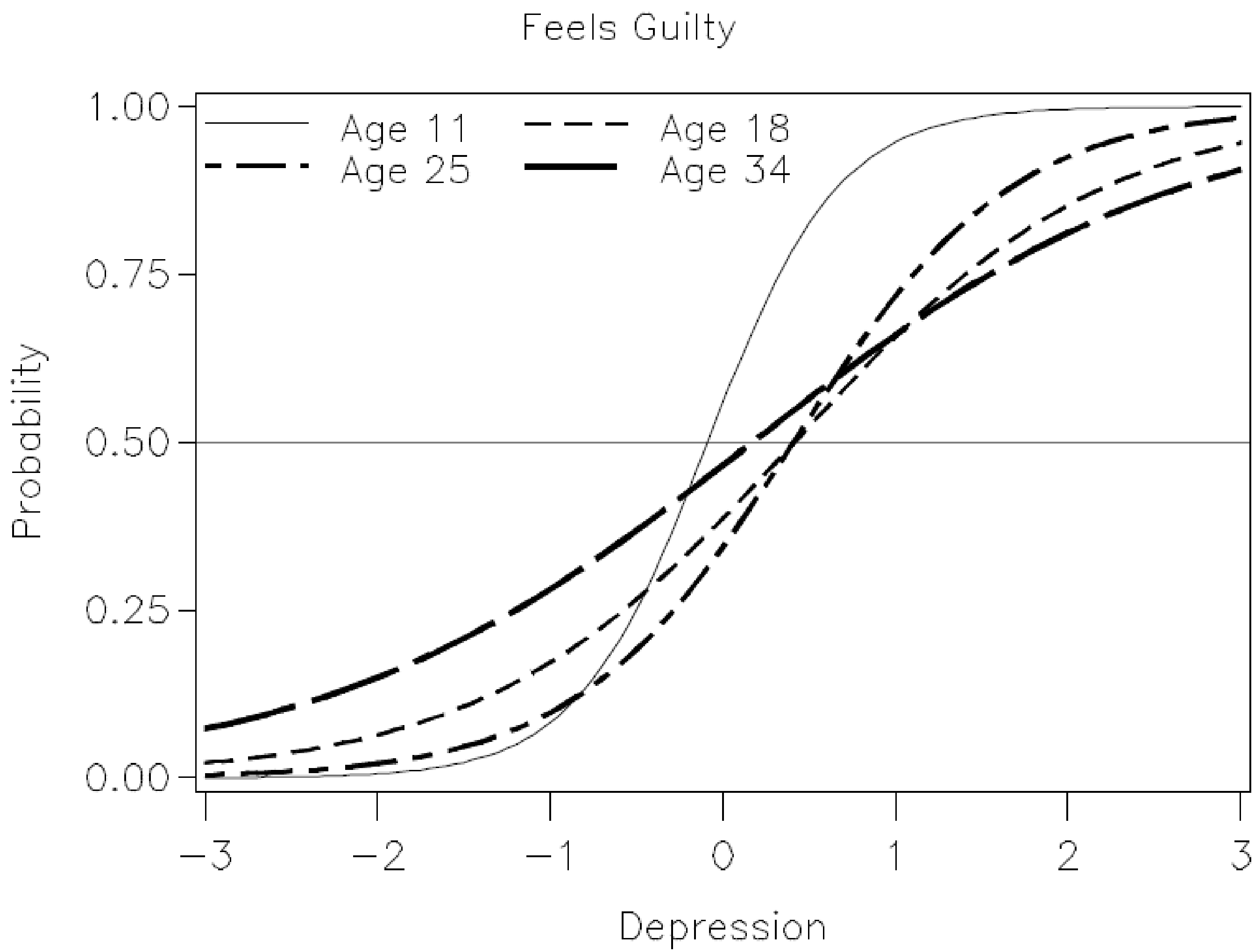
Covariate Effects on Factor

Covariate Effect	Estimate (*SE*)	*t*	*p*
Factor mean			
1. Age	−0.70 (.14)	−5.04	< .0001
2. Age^2	−0.63 (.25)	−2.58	.0098
3. Age^3	0.56 (.18)	3.12	.0018
4. MLS	−0.86 (.08)	−10.87	< .0001
5. AHBP	−0.28 (.12)	−2.28	.0227
6. Gender	−0.48 (.08)	−5.82	< .0001
7. COA	0.23 (.06)	4.07	< .0001
8. Age by MLS	0.85 (.15)	5.81	< .0001
9. Age by AHBP	−3.77 (.91)	−4.13	< .0001
10. Age^2 by AHBP	4.84 (1.64)	2.95	.0033
11. Age^3 by AHBP	−1.73 (.76)	−2.29	.0222
12. Age by Gender	0.25 (.16)	1.58	.1140
13. Age^2 by Gender	1.03 (.31)	3.29	.0010
14. Age^3 by Gender	−0.74 (.23)	−3.18	.0015
Factor variance			
15. Age	0.02 (.10)	0.21	.8372
16. AHBP	−0.13 (.19)	−0.66	.5106
17. Age by AHBP	0.80 (.27)	2.90	.0038

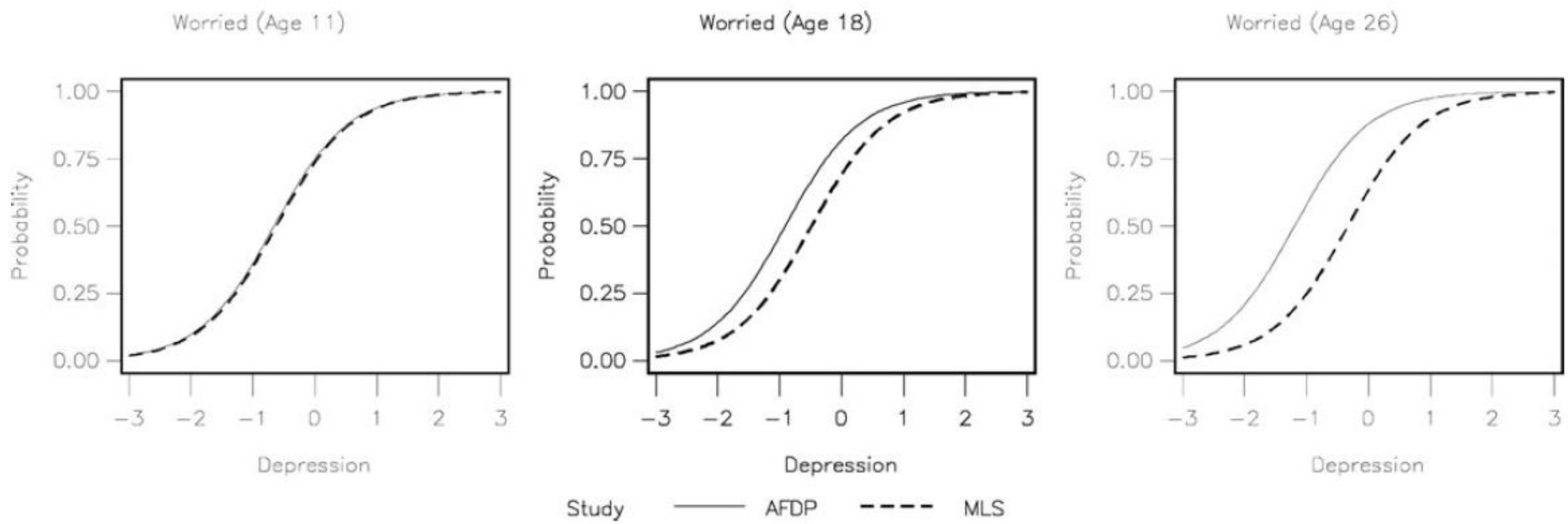
Covariate Effects on Items

Item *Covariate Effect*	Intercept *(SE)*	Loading *(SE)*	Item *Covariate Effect*	Intercept *(SE)*	Loading *(SE)*
1. Lonely	0.79 (.20)	2.36 (.16)	11. Unhappy/Sad/Depressed	0.98 (.19)	1.93 (.19)
AHBP	1.05 (.22)	—	*MLS*	—	0.95 (.29)
2. Cries a lot	0.45 (.15)	1.45 (.12)	12. Worried	1.51 (.18)	1.66 (.13)
Age	−0.34 (.12)	—	*Age*	0.59 (.15)	—
Gender	−2.09 (.17)	—	*MLS*	−0.70 (.18)	—
3. Fears will behave badly	−0.68 (.17)	1.22 (.16)	*Age by MLS*	−0.91 (.31)	—
4. Has to be perfect	0.70 (.11)	1.06 (.08)	13. Hopeless about future	−0.49 (.24)	2.07 (.21)
5. No one loves me	−2.00 (.25)	2.55 (.23)	*Age*	1.04 (.36)	—
Age	−0.64 (.20)	—	*Age*^{2}	−1.41 (.37)	—
MLS	0.88 (.27)	—	*Gender*	0.89 (.24)	—
6. Worthless/Inferior	−1.23 (.20)	2.49 (.19)	14. Acts to harm self	−3.06 (.31)	1.66 (.35)
MLS	0.55 (.21)	—	15. Thinks about killing self	−2.38 (.20)	1.67 (.23)
7. Prefers to be alone	0.18 (.15)	0.91 (.12)	16. Blue	2.02 (.38)	2.83 (.36)
8. Feel guilty	−0.46 (.13)	1.11 (.15)	*Age*	3.33 (.78)	2.96 (.77)
Age	−0.63 (.21)	−0.12 (.32)	17. No interest in things	−0.37 (.25)	1.69 (.21)
Age^2	0.52 (.18)	2.05 (.69)	*Gender*	0.56 (.21)	–
Age^3	–	−1.31 (.42)	*COA*	0.68 (.21)	—
MLS	−0.83 (.19)	—	*Age*	—	2.97 (.88)
9. Is secretive	1.34 (.21)	1.38 (.15)	*Age*^{2}	—	3.02 (.99)
10. Is underactive	0.14 (.16)	1.02 (.13)	*Age*^{3}	—	−2.96 (.93)
Age	0.71 (.24)	—			

DIF Item: *Feels Guilty* by Age



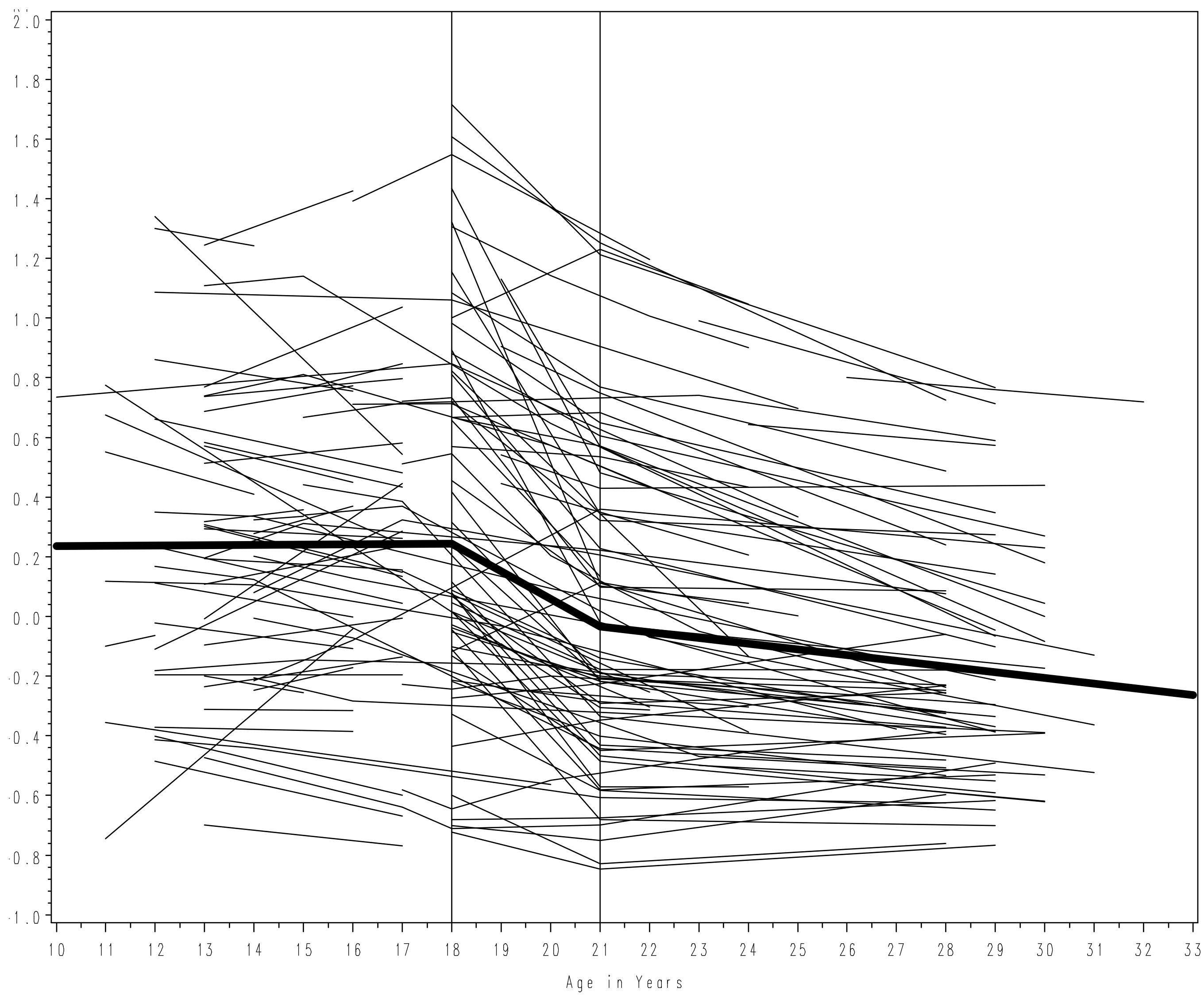
DIF Item: *Worried* by Study & Age



Scoring Phase

- Establish final MNLFA model
  - some items are invariant over covariates
  - some items differentially relate to latent factor as function of covariates
  - latent factor itself differentially relates to covariates
- Take all parameters from final MNLFA and use to obtain optimal scores on depression
  - called Empirical Bayes Estimates of underlying latent factor
- Each subject gets person- and time-specific score of depression that reflects item responses and covariates
  - literally using MNLFA as an incredibly complex calculator

Growth Models Fitted to Scores



Brief Example #2: Substance Use

- Substance use notoriously difficult to measure
  - recall bias, especially for heavy users
  - choosing proper time frame
  - assessing not just consumption but patterns of use
- Challenges particularly salient in children
  - highly episodic
  - low base rates
- Poly-substance use vs. substance-specific use
  - alcohol can hijack polysubstance use

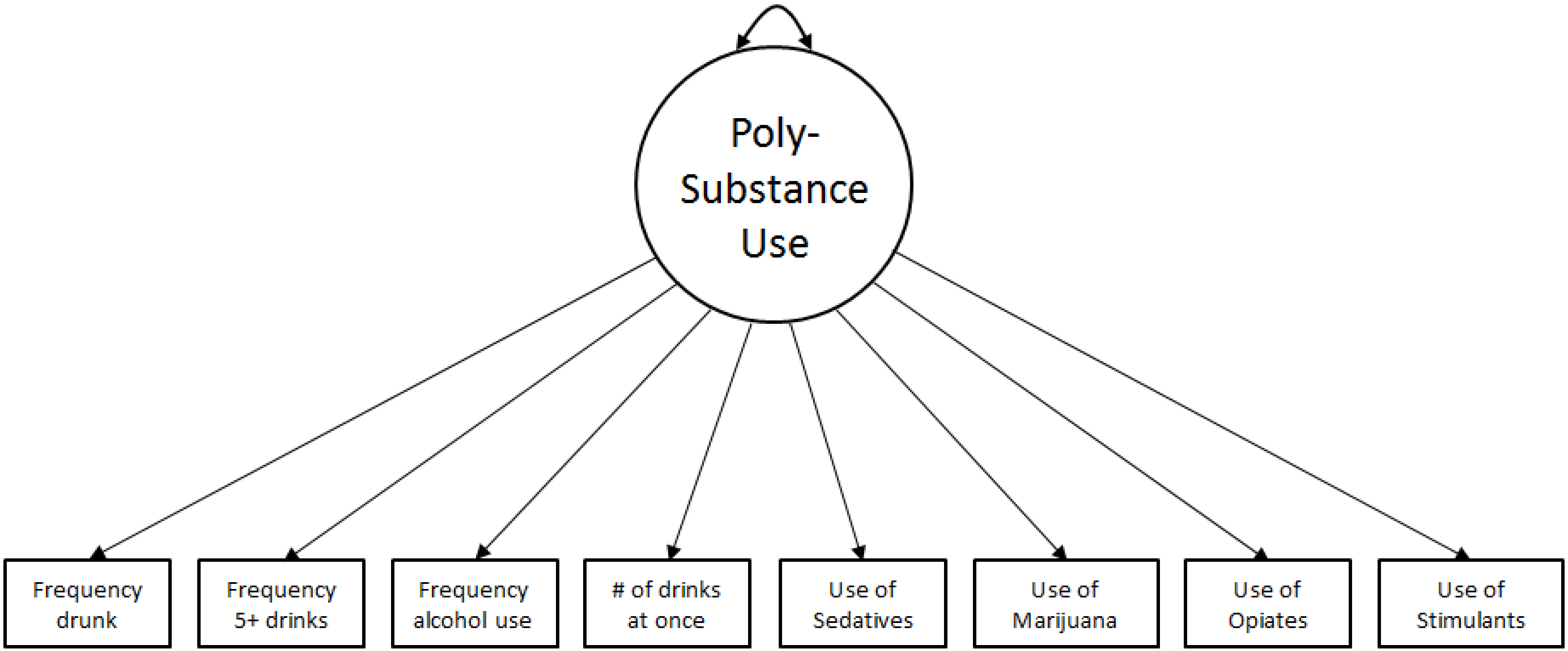
Building a Polysubstance Use Model

- Pool data from 3 studies spanning ages 11 to 35
- Identify 8 drug & alcohol use items
- Define latent variable model of polysubstance use
- Account for alcohol use sub-factor
- Allow for mixture of discrete response scale types
- Assess impact & DIF as function of covariates
- Obtain individual- and age-specific factor score estimates for subsequent analysis

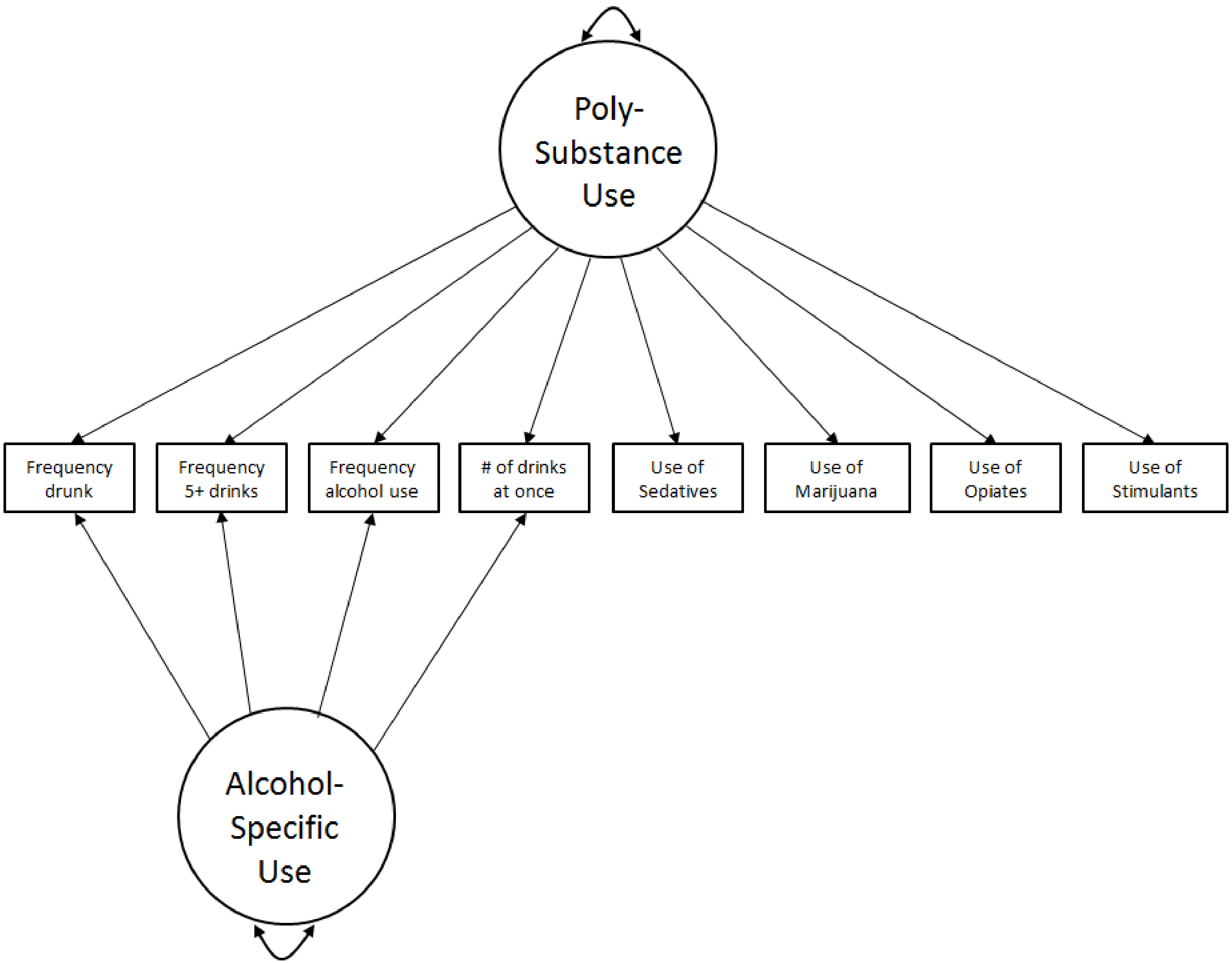
Eight Items from Three Studies

Item	*Item Wording*	*Response Scale*	*Study*
1	Frequency of drunkenness in the past year	0. never 1. less than 6 times a year 2. less than weekly 3. weekly or more often	1 & 2
2	Frequency of drinking 5 or more drinks at one time in the past year	0. never 1. less than 6 times a year 2. less than weekly 3. weekly or more often	2 & 3
3	Frequency of alcohol use (beer, wine, & liquor) in the past year	0. never 1. 1-3 times a month 2. weekly or more often	1, 2 & 3
4	Number of drinks on one occasion in the past year	0. Less than one total drink 1. 1-2 total drinks 2. 3-5 total drinks 3. 6 or more total drinks	2 & 3
5	Use of marijuana in the past year	0. never 1. one or more times	1, 2 & 3
6	Use of stimulants in the past year	0. never 1. one or more times	1, 2 & 3
7	Use of sedatives in the past year	0. never 1. one or more times	1, 2 & 3
8	Use of opiates or hallucinogens in the past year	0. never 1. one or more times	1, 2 & 3

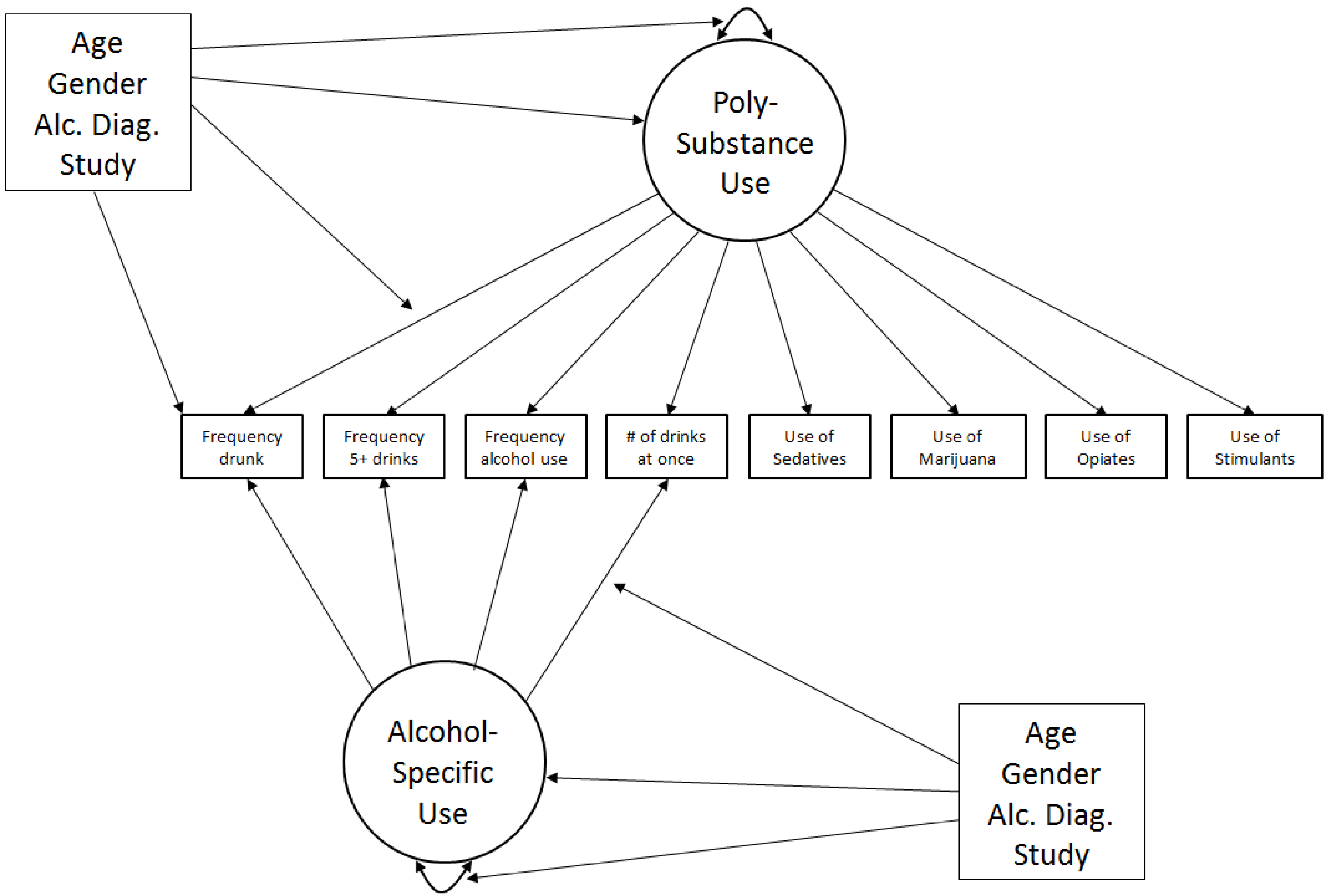
One Factor CFA



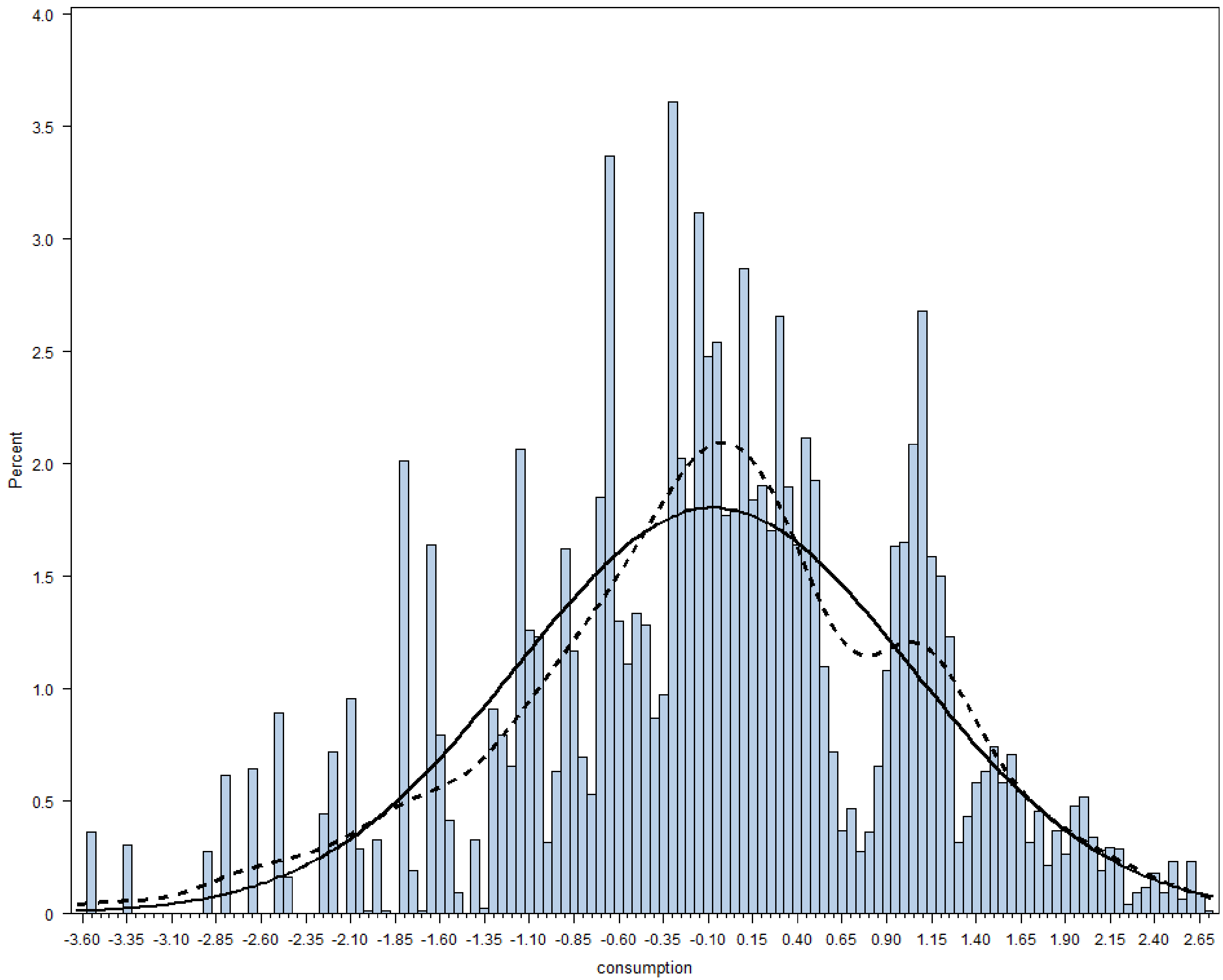
Bifactor CFA



Full Bifactor MNLFA

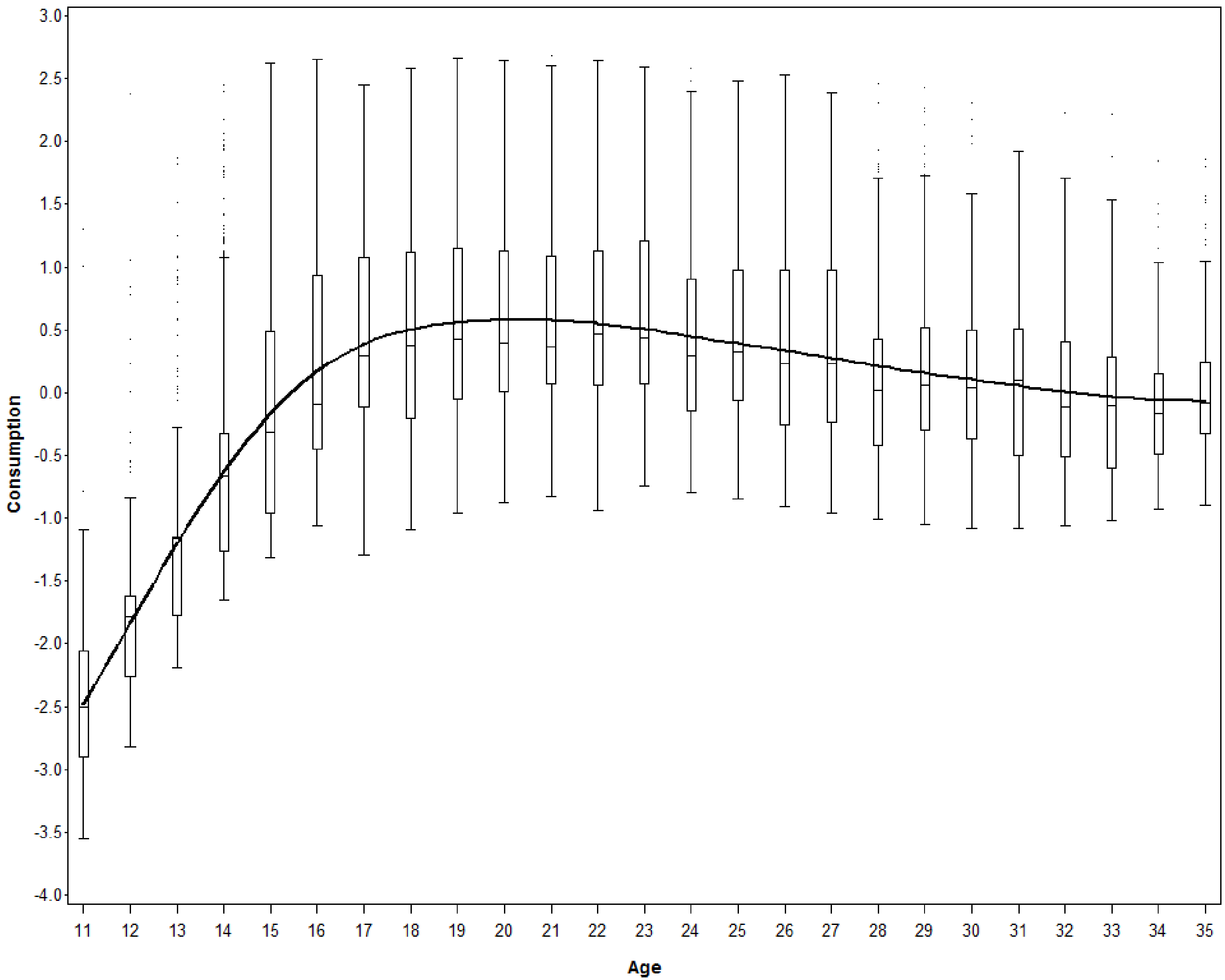


Individual-Specific Scores



Response Scale
0. never 1. less than 6 times a year 2. less than weekly 3. weekly or more often
0. never 1. less than 6 times a year 2. less than weekly 3. weekly or more often
0. never 1. 1-3 times a month 2. weekly or more often
0. Less than one total drink 1. 1-2 total drinks 2. 3-5 total drinks 3. 6 or more total drinks
0. never 1. one or more times
0. never 1. one or more times
0. never 1. one or more times
0. never 1. one or more times

Scores Available for Modeling



Future Directions for IDA

- Our current NIDA-funded project uses true experimental design to validate harmonization procedures and calculation of commensurate measures
- Lab analogue study: REAL U
- Monte Carlo computer simulation study: FAKE U

Future Directions for IDA

- Design of *bridging studies* to link multiple data sets
- Harmonization of discrete diagnostic status measures
  - e.g., linking DSM-IIIR to DSM-IV to DSM-V
- Statistical matching to create synthesized cases
  - current IDA expands data as "long"
  - powerful advantages to expanding data as "wide"
- Develop strategies for designing new data collection efforts in anticipation of future IDA
- All of these extensions need novel development, rigorous evaluation, and broad training

Thank you!