Multilevel Latent Class Analysis (MLCA) to Model Complex Behavior in Intensive Longitudinal Data

Stephanie T. Lanza, Bethany C. Bray, & Yuqi Shen

Society for Behavioral Medicine
Wednesday, April 22, 2026
11 a.m. - 1 p.m. CST

Agenda

11:00-11:15 Refresher on latent class analysis (LCA) and latent profile analysis (LPA)

11:15-12:15 Introduction to MLCA for intensive longitudinal data (e.g., daily diary data, ecological momentary assessments)

- marginal modeling approach
- "nonparametric" multilevel approach

12:15-12:30 Software orientation: Mplus and R
Online resources: Latent Class Analysis Knowledge Base (LCAKB)

12:30-12:40 Demonstration of the multilevLCA R package

12:40-1:00 Q&A

Suggested References:
Multilevel Latent Class Analysis (MLCA) to Model Complex Behavior in Intensive Longitudinal Data

Stephanie T. Lanza, Bethany C. Bray, & Yuqi Shen

Society for Behavioral Medicine
Wednesday, April 22, 2026

Technical papers

Henry, K. L., & Muthén, B. (2010). Multilevel Latent Class Analysis: An Application of Adolescent Smoking Typologies with Individual and Contextual Predictors. *Structural Equation Modeling : A Multidisciplinary Journal*, *17*(2), 193–215. https://doi.org/10.1080/10705511003659342 (Note: Mplus syntax out of date)

Lyrvall, J., Di Mari, R., Bakk, Z., Oser, J., & Kuha, J. (2025). Multilevel Latent Class Analysis: State-of-the-Art Methodologies and Their Implementation in the R Package multilevLCA. *Multivariate Behavioral Research*, *60*(4), 731–747. https://doi.org/10.1080/00273171.2025.2473935

Empirical papers: Marginal approach applied to intensive longitudinal data

Stull, S. W., Marsch, L. A., & Lanza, S. T. (2025). Daily profiles of nondrug reward and their association with daily outcomes for people in recovery from opioid use disorder: An application of latent profile analysis applied to intensive longitudinal data. *Drug and Alcohol Dependence*, *271*, 112646. https://doi.org/10.1016/j.drugalcdep.2025.112646

Linden-Carmichael, A. N., Van Doren, N., Bray, B. C., Jackson, K. M., & Lanza, S. T. (2022). Stress and Affect as Daily Risk Factors for Substance Use Patterns: An Application of Latent Class Analysis for Daily Diary Data. *Prevention Science*, *23*(4), 598–607. https://doi.org/10.1007/S11121-021-01305-9/TABLES/5

Evans-Polce, R. J., Arterberry, B. J., Lanza, S. T., & Patrick, M. E. (2025). Patterns of substance use on a given day in a national sample of U.S. young adults. *Addictive Behaviors*, *168*, 108376. https://doi.org/10.1016/j.addbeh.2025.108376

Empirical papers: Nonparametric approaches in the context of nested data (not ILD yet!)

Halladay, J., MacKillop, J., Munn, C., Amlung, M., & Georgiades, K. (2022). Individual- and school-level patterns of substance use and mental health symptoms in a population-based sample of secondary students: A multilevel latent profile analysis. *Drug and Alcohol Dependence, 240*, 109647. DOI: 10.1016/j.drugalcdep.2022.109647

Mayworm, A. M., Sharkey, J. D., & Nylund-Gibson, K. (2023). An exploration of the authoritative school climate construct using multilevel latent class analysis. *Contemporary School Psychology, 27*, 283-302. DOI:10.1007/s40688-021-00386-1

Vermunt, J. K. (2008). Latent class and finite mixture modeling for multilevel datasets. *Statistical Methods in Medical Research, 17*, 33-51. DOI: 10.1177/0962280207081238

Multilevel Latent Class Analysis (MLCA) to Model Complex Behavior in Intensive Longitudinal Data

Society for Behavioral Medicine

April 22, 2026

Stephanie T. Lanza, Ph.D.

Professor, Biobehavioral Health

The Pennsylvania State University

PennState

Bethany C. Bray, Ph.D.

Interim Assistant Dean for Research, School of Public Health

Associate Director for Scientific and Infrastructure Development, Institute for Health Research and Policy

University of Illinois Chicago

Yuqi Shen, M.H.S.

Doctoral Candidate, Biobehavioral Health

The Pennsylvania State University

PennState

Funding Acknowledgement

NIDA R01-DA057588 (PI: Lanza)

Promoting Rapid Uptake of Multilevel Latent Class Modeling via Best Practices: Investigating Heterogeneity in Daily Substance Use Patterns

A great team!

- Dr. Stephanie Lanza (PI, Penn State)
- Dr. Bethany Bray (Co-I, UIC)
- Dr. Kristina Jackson (Co-I, Rutgers)
- Dr. Ashley Linden-Carmichael (Co-I, U of Oregon)
- Dr. Samuel Stull (Dartmouth)
- Dr. John Dziak (Michigan)
- Ms. Yuqi Shen (Penn State)
- Mr. Danny Wang (Penn State)
- Mr. Sandesh Bhandari (Penn State)

Workshop outline

- Refresher: LCA & LPA
- Introduction to MLCA
- 2 recommended approaches to MLCA
- Software & resources
- R demonstration
- Questions/Discussion

Workshop outline

- **Refresher: LCA & LPA**
- Introduction to MLCA
- 2 recommended approaches to MLCA
- Software & resources
- R demonstration
- Questions/Discussion

Abbreviations

LCA = latent class analysis

- Static, categorical latent variable measured with categorical items

LPA = latent profile analysis

- Static, categorical latent variable measured with continuous items

MLCA/MLPA = multilevel latent class/profile analysis

ILD = intensive longitudinal data

The basic Ideas

- Individuals can be divided into subgroups based on unobservable construct
  - That construct is the latent class variable
  - Subgroups = latent classes
- True class membership is unknown
  - Measurement of the construct is typically based on several indicators

    Categorical indicators: LCA

    Continuous indicators: LPA
- Latent classes are mutually exclusive & exhaustive

An empirical example: Drinking in 12th grade

Data from 2004 Monitoring the Future cohort

n = 2490 high school seniors who answered at least one question about alcohol use (48% boys, 52% girls)

Study goals:

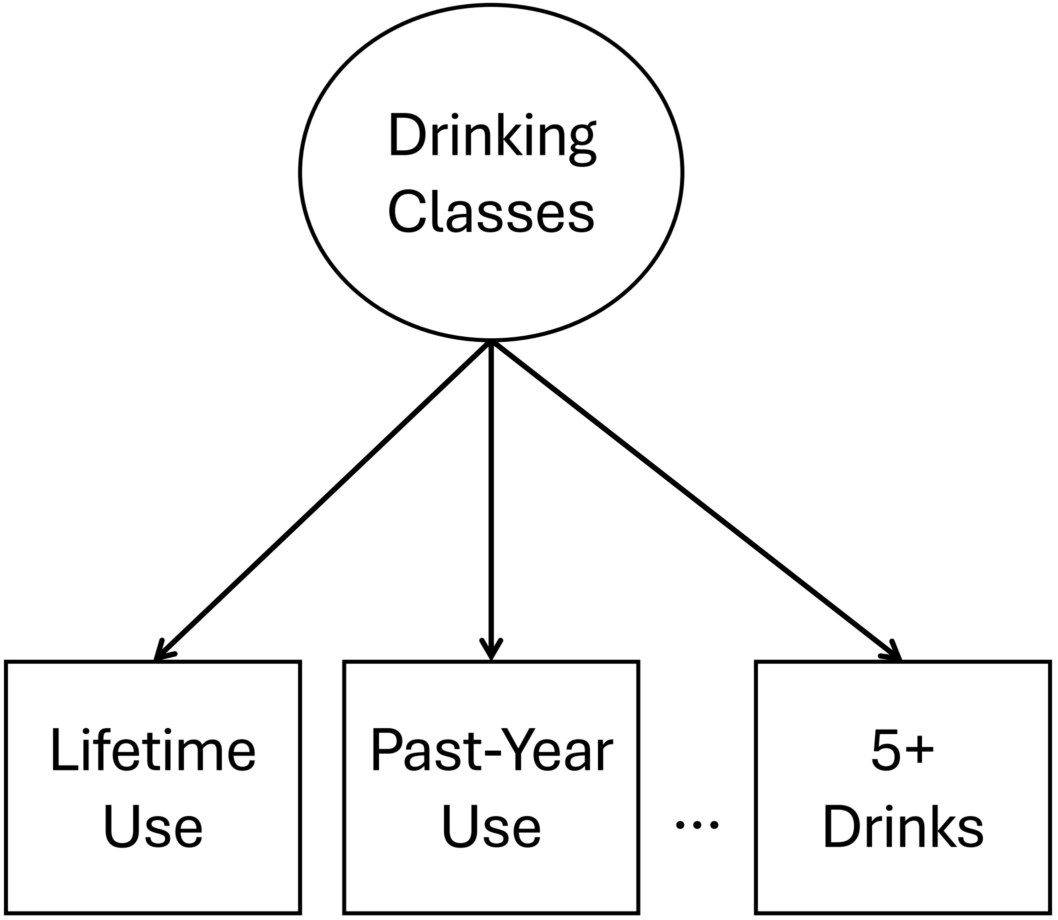
1. Model patterns of alcohol use behavior among U.S. students in Grade 12
2. Predict behavior classes from skipped school (yes/no)

Drinking in 12th Grade

Seven observed indicators of drinking behavior

Item	Proportion 'Yes'
Lifetime alcohol use	82%
Past-year alcohol use	73%
Past-month alcohol use	50%
Lifetime drunkenness	57%
Past-year drunkenness	49%
Past-month drunkenness	29%
5+ drinks in past 2 weeks	26%

LCA/LPA graphical representation



LCA:
Parameters estimated

Indicators are **categorical**!

- **Latent class prevalences**
  - e.g., probability of membership in HEAVY DRINKERS latent class
- **Item-response probabilities**
  - e.g., probability of reporting HEAVY EPISODIC DRINKING given membership in HEAVY DRINKERS latent class

LPA:
Parameters estimated

Indicators are **continuous**!

- **Latent class prevalences**
  - e.g., probability of membership in PROTOTYPICAL leader class
- **Class-specific means**
  - e.g., mean of SINCERE indicator given membership in PROTOTYPICAL leader class
- **Class-specific variances**
  - e.g., variance of SINCERE indicator given membership in PROTOTYPICAL leader class

Latent classes of drinking: 5-class model

Item	Class 1: **Non-drinkers** (18%)	Class 2: **Experi-menters** (22%)	Class 3: **Light Drinkers** (9%)	Class 4: **Past Partiers** (17%)	Class 5: **Heavy Drinkers** (34%)
Lifetime alcohol use	.00	1.00	1.00	1.00	1.00
Past-year alcohol	.00	.61	1.00	1.00	1.00
Past-month alcohol	.00	.00	1.00	.39	1.00
Lifetime drunk	.00	.24	.29	1.00	1.00
Past-year drunk	.00	.00	.00	1.00	1.00
Past-month drunk	.00	.00	.00	.00	.92
5+ drinks past 2 wk	.00	.00	.16	.00	.73

Key assumptions

- LCA & LPA: Indicators are independent within classes (**conditional independence**)
- LPA: Indicators are continuous and normally distributed within classes

Note: MLE typically requires indicator variances to be restricted to be equal across profiles

2 key technical considerations

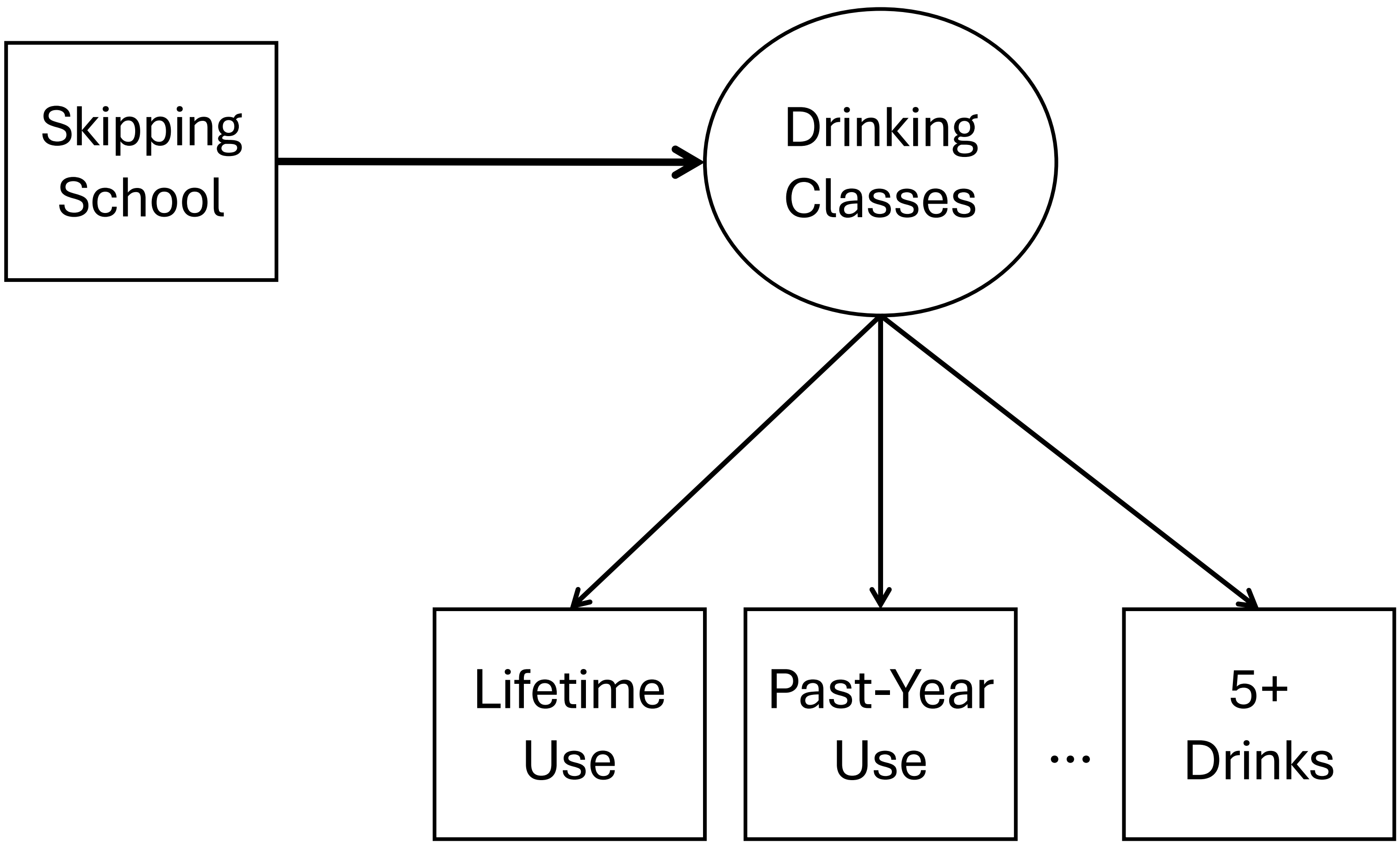
1. Model identification

- Starting values required to “kick off” estimation
- Run many sets of starting values to assess identification

2. Model selection

- Is Model A or Model B better? (relative model fit, e.g., 4 vs. 5 classes)
- **Prior** to adding covariates
- Tools for model selection
  - AIC, BIC, BLRT
  - Latent class separation, interpretation

LCA/LPA with covariates: Graphical representation



LCA/LPA with covariates

- Regress latent class variable on predictors

- Multinomial logistic regression, latent outcome
  - Specify one class as reference class for prediction

- Overall significance test of association
  - Skipped school: $2\,logL$ (4 *df*) = 162.1, $p<.001$

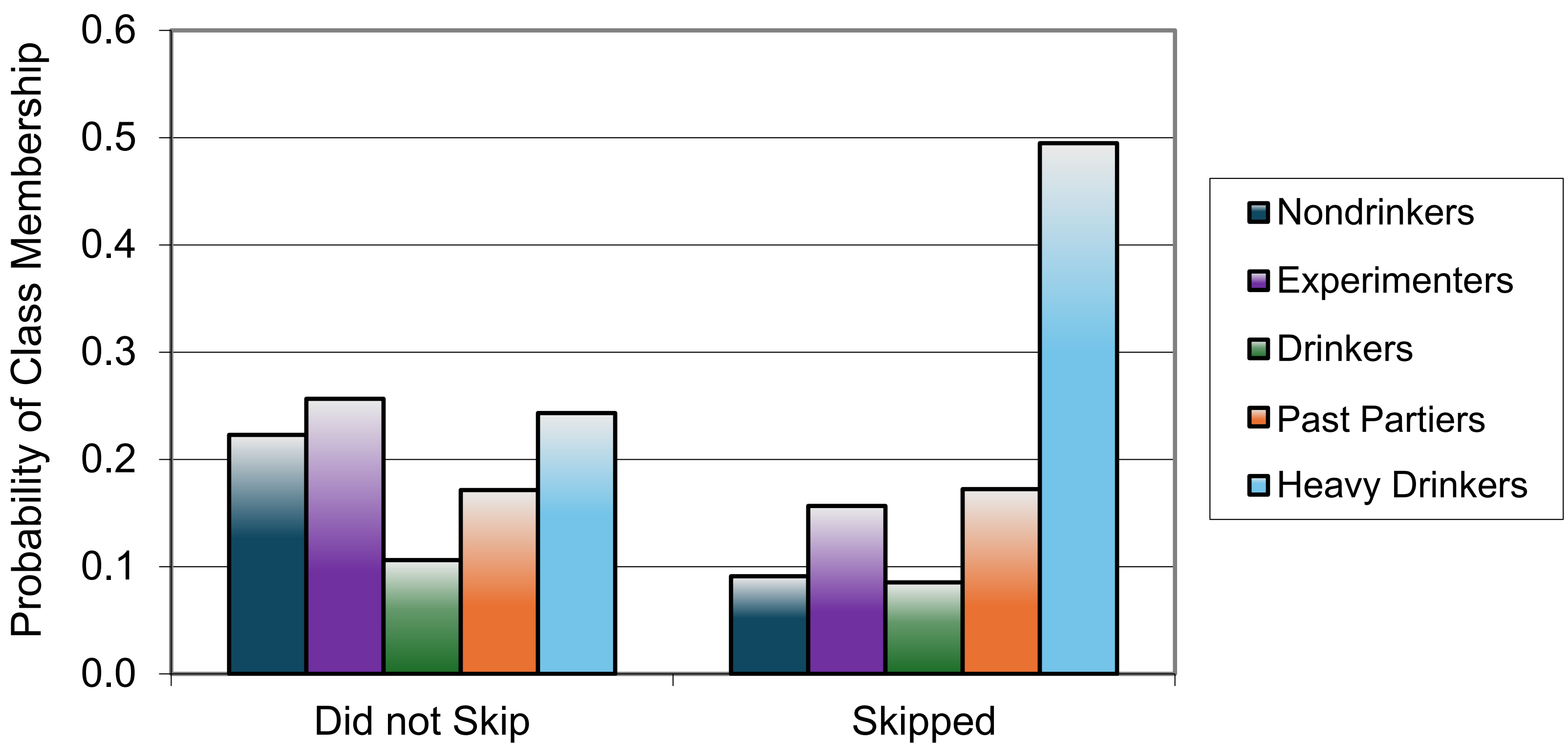
- **Odds ratios** express relation between covariates and class membership

Drinking latent class as function of skipped school

Odds of membership in Heavy Drinkers class relative to the Non-Drinkers class is 5 times higher for adolescents who skipped school relative to those who did not skip.

Class	β	OR
Nondrinkers	---	1.0
Experimenters	0.4	1.5
Light Drinkers	0.7	2.0
Past Partiers	0.9	2.5
Heavy Drinkers	1.6	5.0

Relation between skipped school and drinking classes



Workshop outline

- Refresher: LCA & LPA
- **Introduction to MLCA**
- 2 recommended approaches to MLCA
- Software & resources
- R demonstration
- Questions/Discussion

Many studies have nested data!

- Students nested within classrooms
  - **Level 1:** Student outcome measure, nested within
  - **Level 2:** Classrooms

- Repeated measures (including intensive longitudinal data)
  - **Level 1:** Repeated outcome measure, nested within
  - **Level 2:** Individuals

Remember the main point about nested data

- Defining characteristic: **lack of independence**
  - Most statistical models assume observations are independent
- The problem:
  - **Underestimate SEs**
  - Leads to inflated Type I error rate (false positive)

How do we deal with this, in general?

Approach 1: Generalized estimating equations (GEE)

- Nesting is a “nuisance” – simply adjust SEs

Approach 2: Multilevel modeling (MLM)

- Directly model the clustering variance (random effects)

Multilevel modeling (MLM)

MLM is our go-to analysis tool… why?

1. accounts for nested data structure
2. partitions variance on the outcome
3. can predict Level-1 outcome from:
   - Level-1 covariates
   - Level-2 covariates
   - Cross-level interactions
4. outcome can be **continuous, binary, count…**

fixed $$Y_{ij} = b_{0ij} + b_1 X_{1ij} + b_2 X_{2ij}$$

random $$b_{0ij} = b_0 + u_j + e_{ij}$$

time *i*, person *j*

What if our key construct is multidimensional?

We may wish to model complex, dynamic phenotypes:

1. **Daily behavior patterns** (diet, physical activity, substance use)
2. **Affective states** (high-activation PA, NA; low-activation PA, NA)
3. **Momentary contextual risk** (location, people around you, noise)
4. **Daily social media use patterns**

What are other examples?

Multilevel latent class analysis (MLCA)

MLCA applied to ILD is a framework for modeling

- Dynamic, multidimensional constructs
- Moment-level covariates of moment-level class membership
- Person-level covariates

Latent class variable: **Level-1 outcome**

Probability of membership in latent classes can vary across **moments/days** and across **people**

Workshop outline

- Refresher: LCA & LPA
- Introduction to MLCA
- **2 recommended approaches to MLCA**
- Software & resources
- R demonstration
- Questions/Discussion

Where do I begin?

First and foremost:
Focus on the measurement model for your construct

- What dynamic construct am I trying to measure?
- What repeatedly measured variables will I include as indicators?
- No covariates yet!
- Using all available data:
  - Confirm model identification
  - Conduct Level-1 model selection (How many latent classes? Interpretable?)

How do I estimate the model?

Four modeling approaches have been proposed:

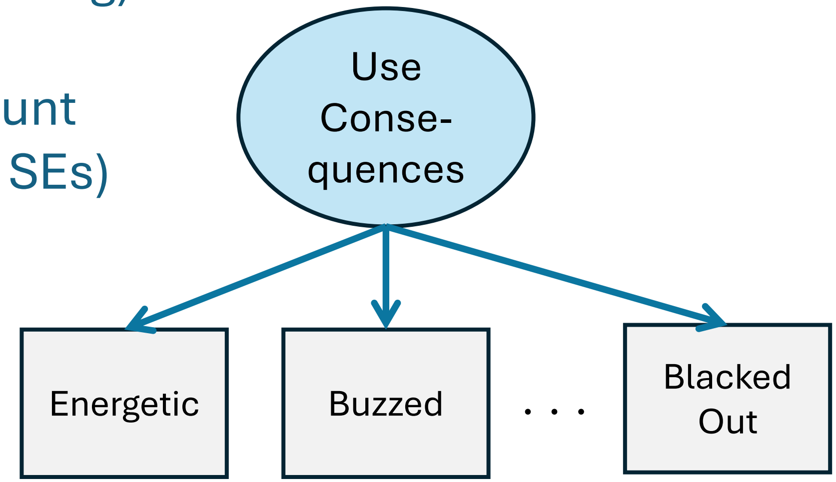
a. Marginal model (GEE analogue)
b. Random effects model (MLM equivalent)
c. Continuous factor at Level 2
d. Latent class variable at Level 2 (“nonparametric” approach)

Aren’t options 1 and 2 sufficient?

A graphical view:
a) Marginal model

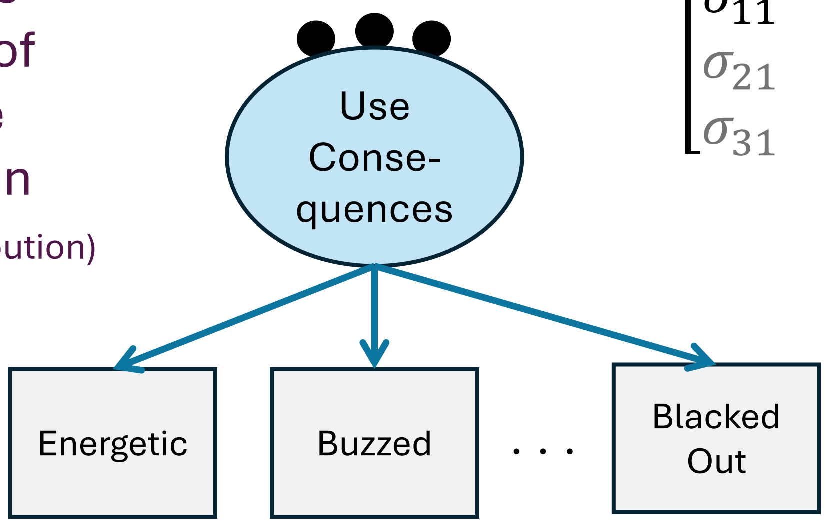
Parameter estimates identical to standard LCA model (ignoring clustering)

SEs adjusted to account for clustering (robust SEs)



A graphical view:
b) Full random effects model

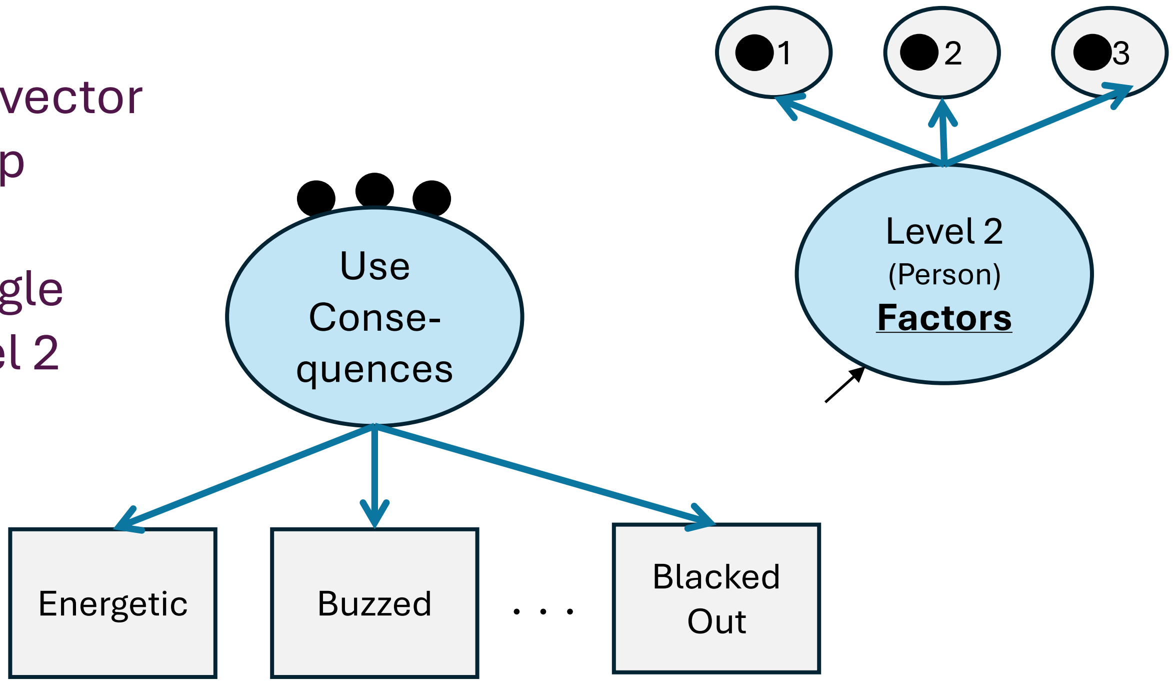
- 4 classes, so 3 random intercepts
- Allow probability of use consequence class to vary within people (normal distribution)
- Random effects correlated or uncorrelated



$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

A graphical view:
c) Use continuous factor at level 2

- 4 classes
- Each individual has vector of class membership probabilities
- Used to indicate single latent **factor** at Level 2

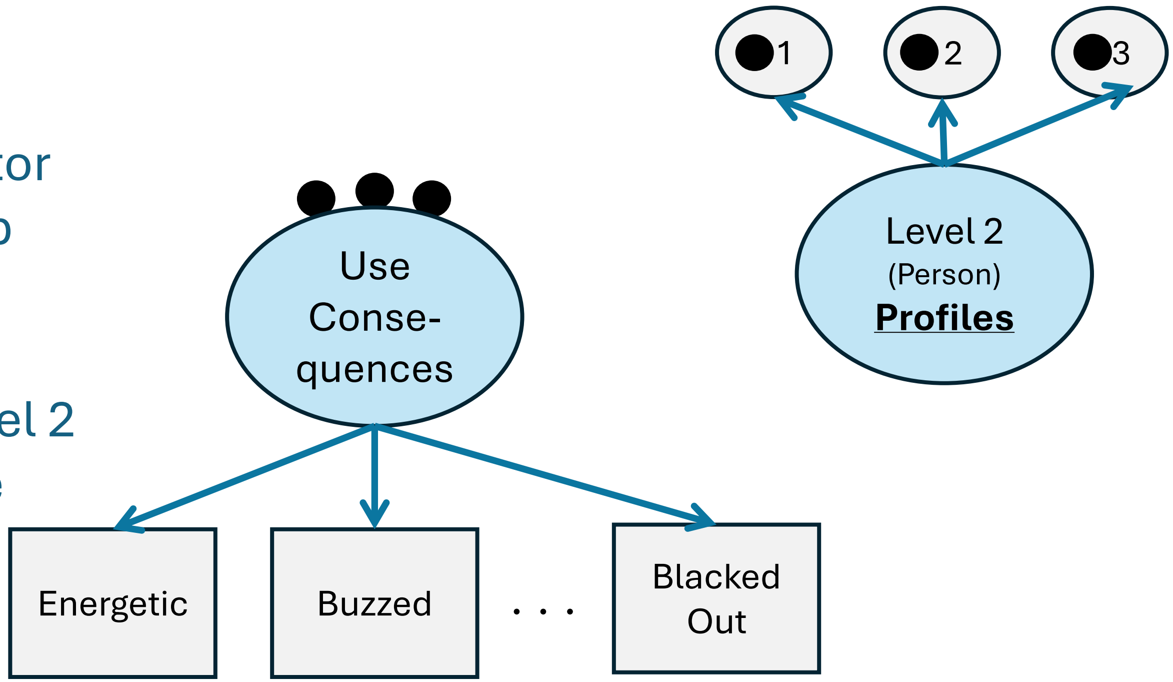


A graphical view:
d) Use latent class variable at Level 2

4 classes

Each individual: vector of class membership probabilities

Used to indicate Level 2 **classes** (sounds like LPA!)



Let’s compare approaches!

Estimation Approach	Relative Fit	Estimation Speed	Model Identification	Interpretation	Separates Variance
Marginal Model					
Random Effects (full multilevel model)					
Factor at Level 2					
Classes at Level 2					

Let’s compare approaches!

Estimation Approach	Relative Fit	Estimation Speed	Model Identification	Interpretation	Separates Variance
Marginal Model	**WORST**	**VERY FAST**	**GOOD**	**EASY**	**NO**
Random Effects (full multilevel model)	**BEST**	**VERY SLOW**	**WORST**	GOOD	**YES**
Factor at Level 2	MODERATE	SLOW	BAD	**VERY DIFFICULT**	YES (SORT OF)
Classes at Level 2	GOOD	FAST	**GOOD**	GOOD	YES (SORT OF)

Let’s compare approaches!

Estimation Approach	Relative Fit	Estimation Speed	Model Identification	Interpretation	Separates Variance
Marginal Model	**WORST**	**VERY FAST**	**GOOD**	**EASY**	**NO**
Random Effects (full multilevel model)	**BEST**	**VERY SLOW**	**WORST**	GOOD	**YES**
Factor at Level 2	MODERATE	SLOW	BAD	**VERY DIFFICULT**	YES (SORT OF)
Classes at Level 2	GOOD	FAST	**GOOD**	GOOD	YES (SORT OF)

“Nonparameteric Approach”

Recommended Approach #1

Marginal Model

(shown with Mplus)

Let's take an example through **marginal approach**

- **21 days** - daily diary study, ~33,000 days
- **7,629 substance use days**
  - mean 6.1 days/person
- **1,266 college students** reported substance use on 1+ days
  - mean age 20.1 years
  - 72% female, 84% NH White

Let's take an example through **marginal approach**

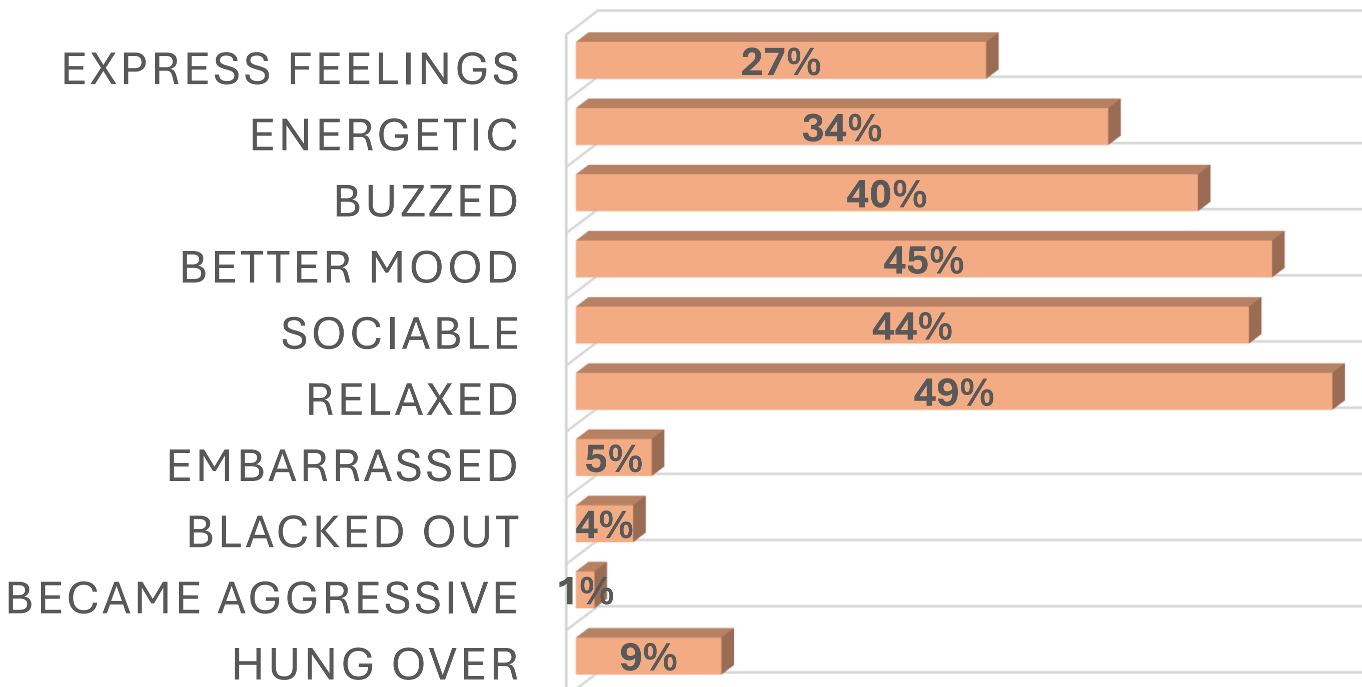
7,629 days (Level 1)
nested within
1,266 students (Level 2)

Outcome:

- Each day's substance use consequence pattern (i.e., day-level latent class)

Standard LCA based on 7,629 substance use days

10 Level-1 indicators: Consequences of today's use (binary)



Standard LCA based on 7,629 substance use days

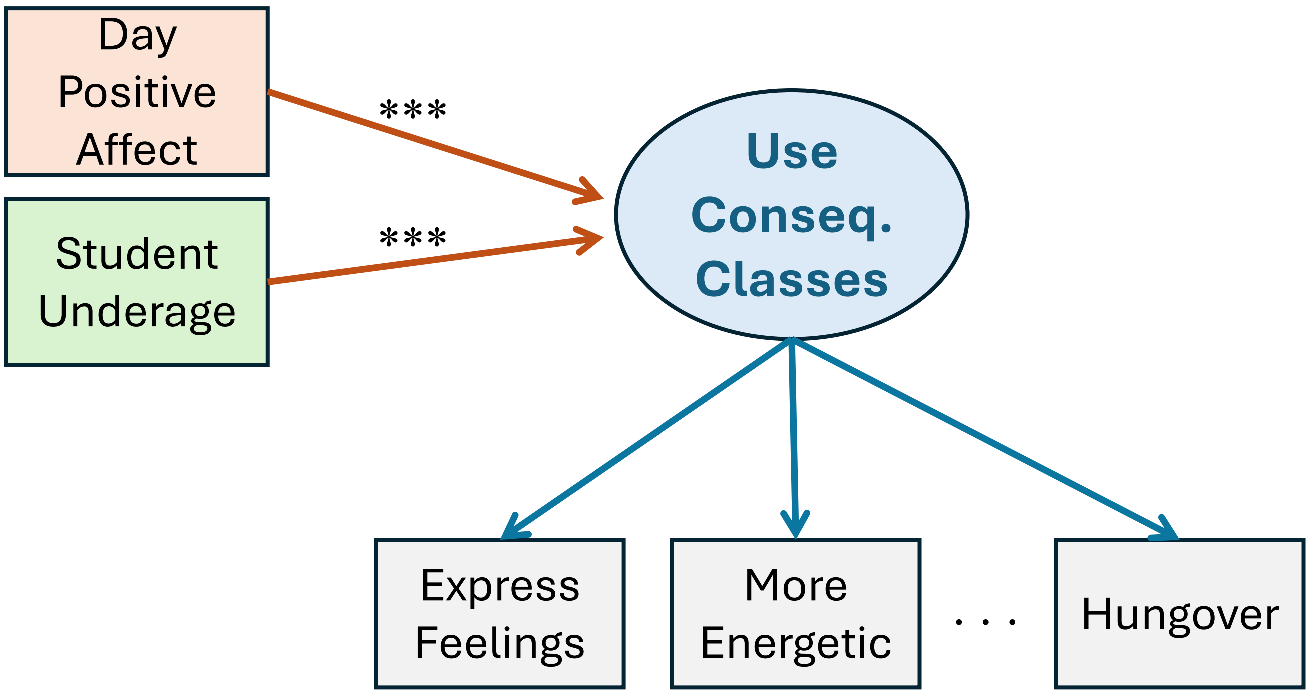
Step 1. Conduct Level 1 model selection (we selected 6 day-level consequence classes)

Consequence:	Broad Positive+ Negative Conseq. (3%)	Positive Social/ Negative Physical (4%)	Energized/ Social (21%)	Broad Positive Conseq. (15%)	No Conseq. (40%)	Happy/ Relaxed (17%)
Express feelings	.90	.38	.32	.79	.03	.16
Energetic	.95	.36	.62	.83	.08	.10
Buzzed	.85	.42	.46	.80	.21	.35
Better mood	.97	.25	.55	.98	.06	.73
Sociable	1.00	.68	.73	1.00	.07	.29
Relaxed	.69	.23	.25	.87	.31	.91
Embarrassed	.55	.44	.01	.05	.00	.01
Blacked out	.45	.31	.00	.03	.00	.01
Became aggressive	.17	.07	.01	.01	.00	.00
Hung over	.67	.57	.10	.10	.03	.01

Step 2. Add cluster statement (Person ID)

- This simply adjusts (i.e., increases) SEs
- Parameter estimates unchanged!

Step 3. Add covariates, interpret ORs
A graphical view:



Step 3. Add covariates, interpret ORs

- Level-1 and Level-2 covariates can be included
- Both predict probability of class on a **day**

Covariate	Broad Positive+ Negative Conseq.	Positive Social/ Negative Physical	Energized/ Social	Broad Positive Conseq.	No Conseq.	Happy/ Relaxed
Daily positive affect (2LL=812.67, 5df, p<.0001)	**2.48***	**1.38***	**1.63***	**1.69***	*ref*	**0.85***
Student underage (2LL=63.30, 5df, p<.0001)	1.57	0.77	1.26	**1.84***	*ref*	1.13

Step 3. Add covariates, interpret ORs

All relative to No Consequences class:

- Higher daily PA→ **lower** odds of Happy/Relaxed
- Higher daily PA → **higher** odds of
  - Broad Positive + Negative
  - Positive Social, Negative Physical
  - Energized/Social
  - Broad Positive

- Underage student → **higher** odds of Broad Positive Conseq.

- *Largest effect size:* Holding underage constant, every one unit higher on PA corresponds to an average of **2.5 times greater odds** of having a use day with broad positive and negative consequences, across all people and all days

Recommended Approach #2

Nonparametric (i.e., Class at Level 2)

(shown with R)

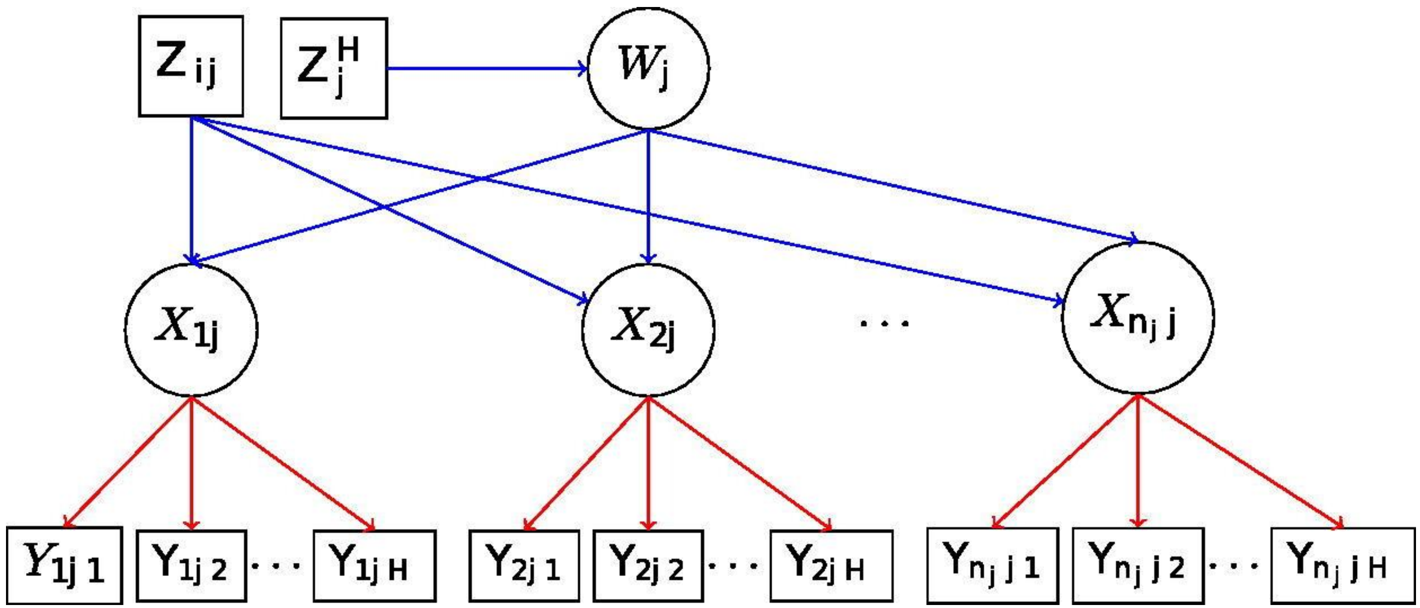
What are we actually fitting??

Let's take example through **nonparametric approach**

i: Level-1 units
X: Level 1 classes

j: Level-2 units
W: Level 2 classes

Z_{ij}: Daily PA
Z_j: Underage



Lyrvall, Mari, Bakk, Oser, & Kuha (2025)

What are we actually fitting??

Let's take example through **nonparametric approach**

$$P(\mathbf{Y}_{ij}|\mathbf{Z}_{ij}) = \sum_{m=1}^{M} P(W_j = m|\mathbf{Z}_j^H)[\sum_{t=1}^{T} P(X_{ij} = t|W_j = m, \mathbf{Z}_{ij}) \prod_{h=1}^{H} P(Y_{ijh}|X_{ij} = t)]$$

The conditional response probabilities $P(Y_{ijh}|X_{ij} = t)$ define the LC *measurement model*, while the conditional class membership probabilities $P(W_j = m|\mathbf{Z}_j^H)$ and $P(X_{ij} = t|W_j = m, \mathbf{Z}_{ij})$ define the LC *structural models*.

Lyrvall, Mari, Bakk, Oser, & Kuha (2025)

Step 1. Select, interpret Level-1 classes (we've seen this before!)

Consequence:	Broad Positive+ Negative Conseq. (3%)	Positive Social/ Negative Physical (4%)	Energized/ Social (21%)	Broad Positive Conseq. (15%)	No Conseq. (40%)	Happy/ Relaxed (17%)
Express feelings	.90	.38	.32	.79	.03	.16
Energetic	.95	.36	.62	.83	.08	.10
Buzzed	.85	.42	.46	.80	.21	.35
Better mood	.97	.25	.55	.98	.06	.73
Sociable	1.00	.68	.73	1.00	.07	.29
Relaxed	.69	.23	.25	.87	.31	.91
Embarrassed	.55	.44	.01	.05	.00	.01
Blacked out	.45	.31	.00	.03	.00	.01
Became aggressive	.17	.07	.01	.01	.00	.00
Hung over	.67	.57	.10	.10	.03	.01

Step 2. Add Level-2 class variable
→ Select, interpret Level-2 classes

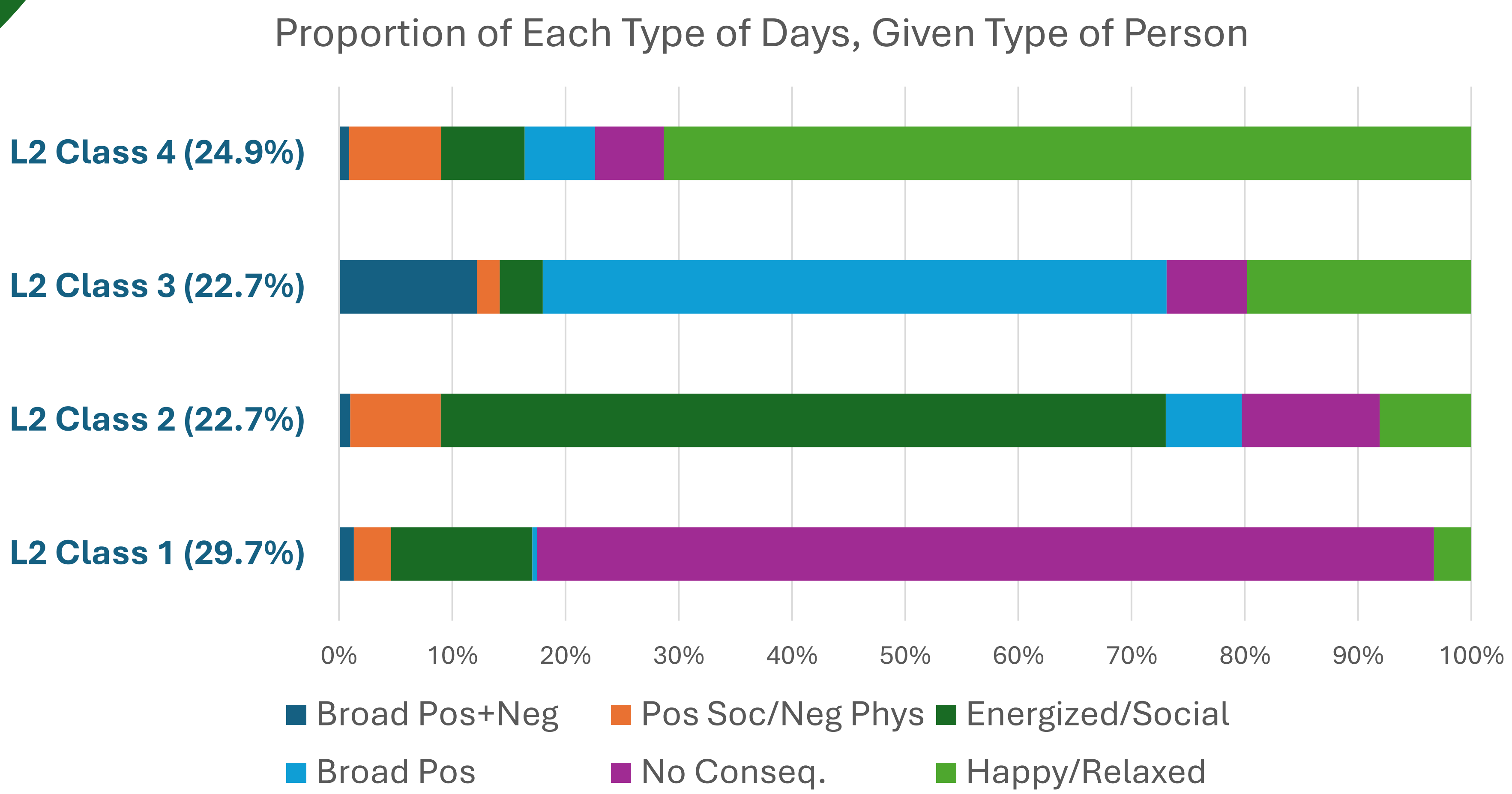
Classes	LogL	Parameters	AIC	BIC	aBIC	Entropy
1	-30817.1	65	61764.1	62215.2	62008.6	0.70
2	-29883.4	71	59908.8	60401.5	60175.9	0.77
3	-29460.9	77	59075.8	59610.2	59365.5	0.78
4	**-29238.8**	**83**	**58643.6**	**59219.6**	**58955.9**	**0.79**
5	-29079.5	89	58337.0	58954.6	58671.8	0.81
6	-28999.9	95	58189.8	58849.1	58547.2	0.80

Remember, we're now doing LCA at Level 1 & LPA at Level 2

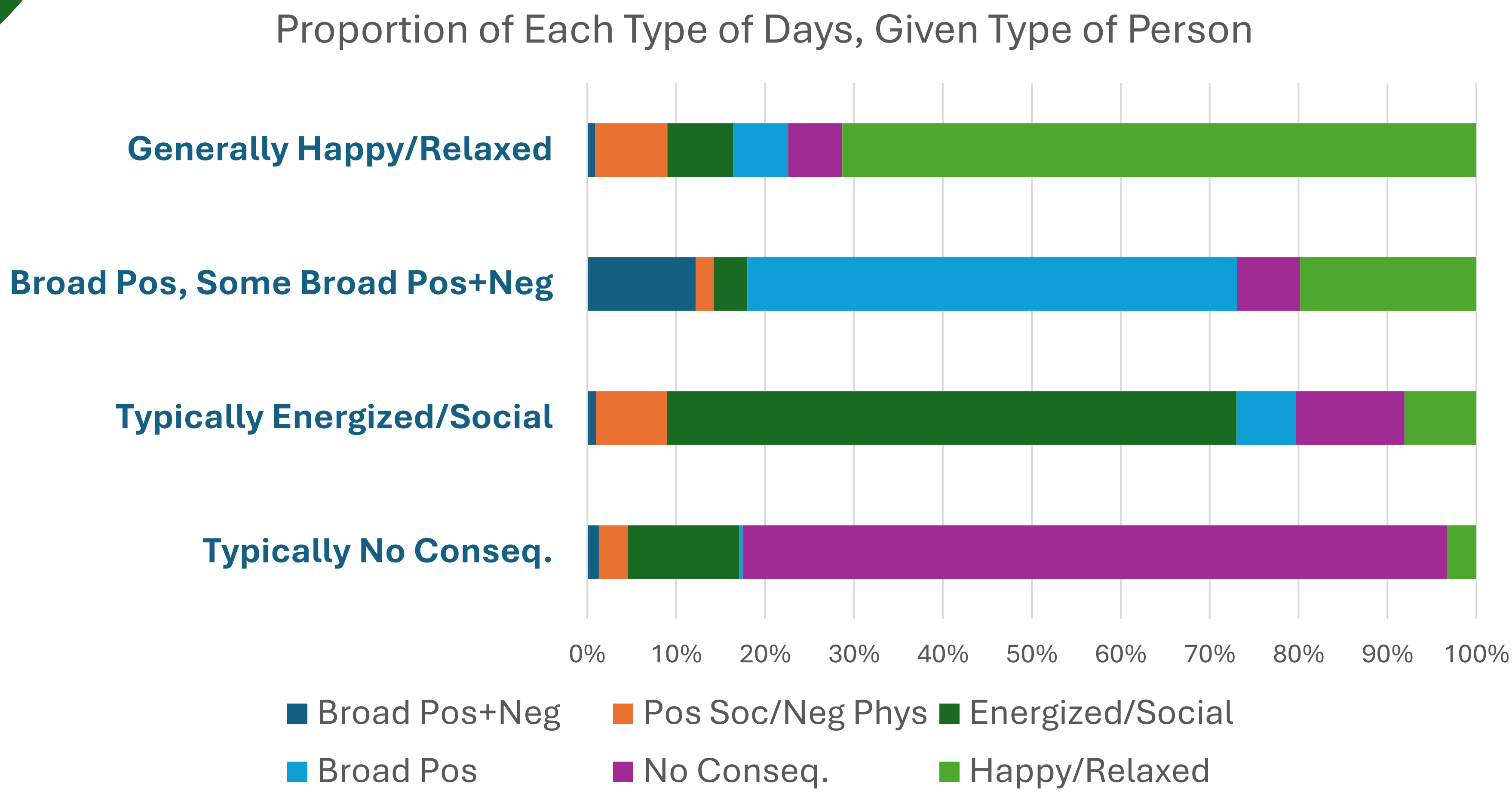
Note that item-response probabilities will be somewhat changed: Check L1 interpretation again

Consequence:	Broad Positive+ Negative Conseq.	Positive Social/ Negative Physical	Energized/ Social	Broad Positive Conseq.	No Conseq.	Happy/ Relaxed
Express feelings	.92	.40	.33	.74	.03	.07
Energetic	.97	.44	.62	.74	.10	.07
Buzzed	.90	.45	.45	.76	.21	.28
Better mood	.99	.33	.58	.97	.04	.42
Sociable	1.00	.75	.73	.94	.07	.17
Relaxed	.77	.24	.23	.88	.07	.85
Embarrassed	.47	.40	.01	.04	.00	.01
Blacked out	.42	.28	.00	.02	.00	.01
Became aggressive	.14	.08	.01	.00	.00	.00
Hung over	.58	.59	.08	.08	.04	.01

Now interpret Level-2 classes (types of students)



Now interpret Level-2 classes (types of students)



Step 3. Add covariates, interpret ORs

- Level-1 and Level-2 covariates can be included
- Both can predict probability of experiencing certain consequence class on a day

- **Level-2 covariates can now predict type of person!**
- **Effect of Level-1 covariates can vary across Level-2 classes!**

Step 3. Add covariates, interpret ORs

Effect of Underage at Level 2:

	Typically No Conseqences	Typically Energized/ Social	Broad Pos, Some Broad Pos+Neg	Generally Happy/Relaxed
Student Underage	*ref*	**1.20***	**1.81***	**0.78***

Step 3. Add covariates, interpret ORs

Effect of Positive Affect at Level 1:

Level-1 Effects Conditional on Level-2 Class	Broad Positive+ Negative Conseq.	Positive Social/ Negative Physical	Energized/ Social	Broad Positive Conseq.	No Conseq.	Happy/ Relaxed
Typically No Conseqences	**2.27***	1.71	2.05	1.07	*ref*	1.42
Typically Energized/Social	**11.29***	2.10	**2.49***	**5.27***	*ref*	1.22
Broad Pos, Some Broad Pos+Neg	**1.61***	1.19	**0.64***	1.07	*ref*	**0.48***
Generally Happy/Relaxed	**4.57***	1.44	**2.12***	2.31	*ref*	0.94

(preliminary) Conclusions, next steps

- ***On outcomes:*** Can model more complex phenotpyes!
- ***On random effects:*** If nested data, we do not necessarily need full random effects model
  - Not plausible with latent class outcome
- ***On variance:*** Only separate variance if you need to
  - Marginal model may be sufficient!
- Much work remains to clarify interpretation, utility, cross-level effects using each approach

Marginal modeling approach can work for many studies

1) Can use Mplus, Latent Gold, SAS (probably R)
2) Extension to predicting outcome at t+1 from Level-1 class membership at t clear

Nonparametric approach:
Shows most promise for complex questions

Step 1. Conduct L1 model selection

- Use standard tools, all available person-moments

Step 2. Specify 2-level model

- Conduct L2 model selection using standard tools
- Interpret L2 classes

Step 3. Add covariates

- L1, L2, or both

Workshop outline

- Refresher: LCA & LPA
- Introduction to MLCA
- 2 recommended approaches to MLCA
- **Software & resources**
- R demonstration
- Questions/Discussion

New software

multilevLCA R package

Lyrvall, Mari, Bakk, Oser, & Kuha (2025) Multivariate Behavioral Research

6 Classes (within), 4 classes (between) without covariates – nonparametric approach

```
Model = multiLCA(data = df, Y = c("con_1", "con_2", "con_3", "con_4",
     "con_5", "con_6", “con_7", "con_8", “con_9", “con_10"),
     iT = 6, id_high = "ID_temp", iM = 4, incomplete = TRUE,
     kmea = TRUE, extout = TRUE, sequential = FALSE, numFreeCores = 8,
     reord = TRUE, fixedpars = 1, verbose = TRUE)
summary(Model)
```

Data structure for LCA

Rectangular

One row per subject

One column for each variable

Can include covariates

Sometimes called "wide" data structure

Subject ID | Indicators | Covariate

Student ID	Lifetime Drink	Past-Yr Drink	Felt Drunk	5+ Drink	Age
1	0	0	0	0	14
2	1	0	0	0	15
3	1	1	1	1	17
4	1	1	0	0	14
5	1	0	1	0	16

and so on…

Data structure for MLCA

Rectangular

One row per subject

One column for each variable

Can include covariates for prediction at **both levels**

Add column indicating person

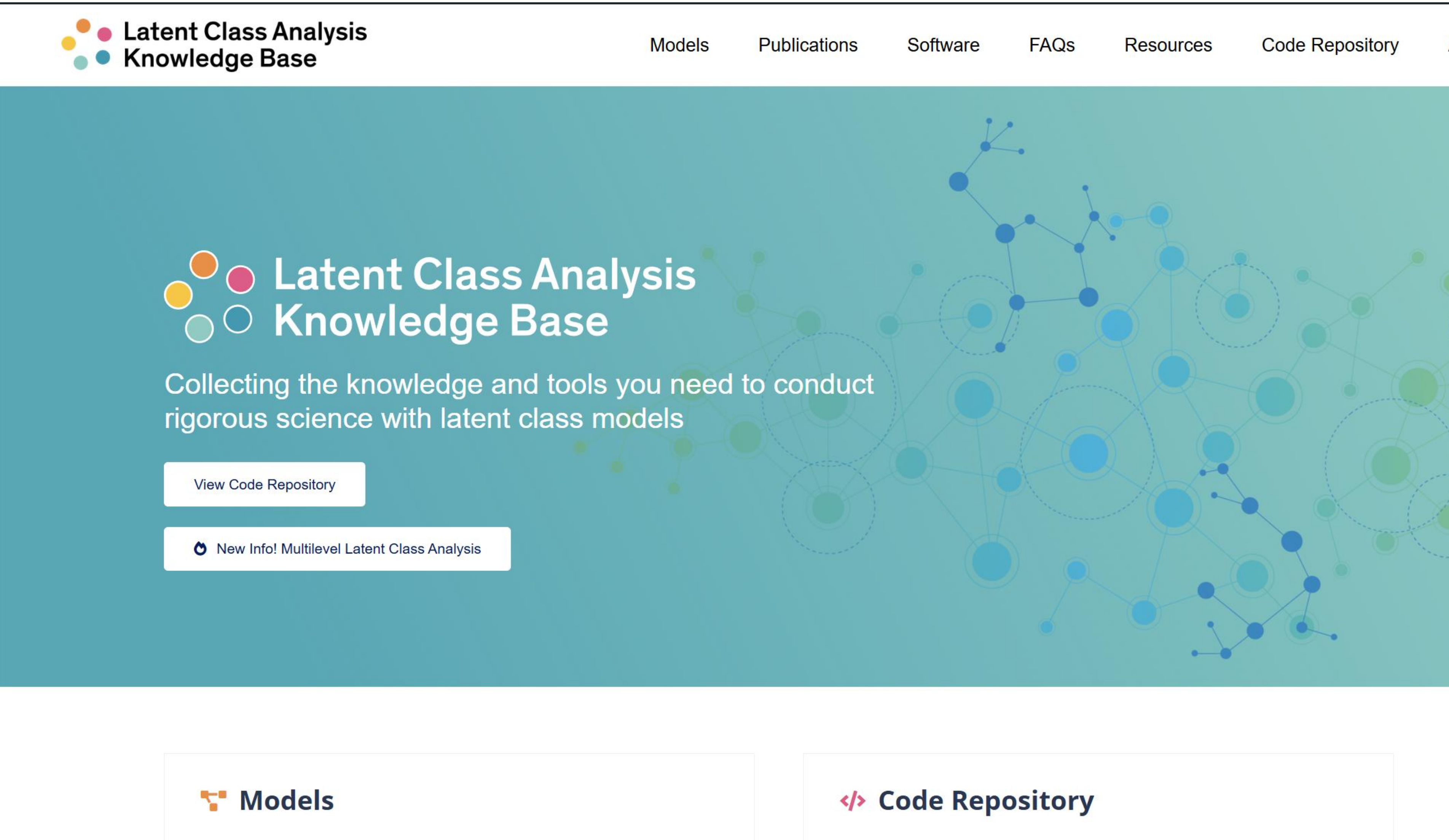
Level-1 ID	Level-2 ID	Indicators				L1 Cov.	L2 Cov.
Time	Person ID	Lifetime Drink	Past-Yr Drink	Felt Drunk	5+ Drink	Affect	Under-age
1	1	0	0	0	0	14	0
2	1	1	0	0	0	15	0
3	1	1	1	1	1	17	0
4	2	1	1	0	0	14	1
5	2	1	0	1	0	16	1

and so on…

Latent Class Analysis Knowledge Base

www.latentclassanalysis.com

Learn more about fitting a variety of latent class models, find sample code, and more!


Latent Class Analysis Knowledge Base
Models
Publications
Software
FAQs
Resources
Code Repository
Latent Class Analysis Knowledge Base
Collecting the knowledge and tools you need to conduct rigorous science with latent class models
View Code Repository
New Info! Multilevel Latent Class Analysis
Models
Code Repository

Multilevel Latent Class Analysis (MLCA)

Description

Multilevel latent class analysis (MLCA) was developed more than a decade ago to address nested data, such as individuals nested within a higher-order structure (e.g., students nested in classrooms). By applying MLCA to the analysis of intensive longitudinal data (ILD), we can better understand the heterogeneity of behavior patterns in daily life and identify within-person vs. between-person risk factors. Specifically, we can model comprehensive, within-day substance use patterns, incorporate day-level psychosocial and person-level predictors of patterns, and predict acute (e.g., next-day) outcomes.

Modeling Latent Class Variables in the Context of Intensive Longitudinal Data (ILD)

Marginal Modeling Approach to MLCA

Marginal models (e.g.., generalized estimating equations applies to longitudinal data), are a popular alternative to random effects models for repeated-measures data. Marginal models estimate average effects in a population and produce cluster-adjusted, robust standard errors;

Two-Level Modeling Approach to MLCA with latent classes at both levels

This approach to random effects modeling employs latent class variables at the day-level and person-level. The person-level classes summarize the random effects of interest and reflect heterogeneity across persons in their probabilities of having certain types of days. In some of the

Workshop outline

- Refresher: LCA & LPA
- Introduction to MLCA
- 2 recommended approaches to MLCA
- Software & resources
- **R demonstration**
- Questions/Discussion

Software demonstration: multilevLCA R package Nonparametric approach with no covariates

```
test_mod_consequence1<-multiLCA(
        data = dt_new,
        #specify the dataset
        Y = c("con_1","con_2","con_3", "con_4","con_5",
                      "con_6","emb_rude","con_9",
                      "agg_hurt","hungover"),
        #specify the indicators
        iT = 6,
        # lower level number of classes
        id_high = "ID_temp",
        # higher level grouping variable
        iM = 4,
        # higher level number of classes
        sequential = FALSE,
        NRmaxit = 10000,
        # Maximum number of iterations for
        # Newton-Raphson algorithm. Default: 100.
        NRtol = 1e-20,
        # Tolerance for Newton-Raphson algorithm.
        # Default: 1e-6.
        maxIter = 1e6,
        # Maximum number of iterations for
        # EM algorithm. Default: 1e3.
        tol = 1e-20,
        # Tolerance for EM algorithm.
        # Default: 1e-8.
        numFreeCores = 8,
        reord = TRUE,
        # customizing reference class
        reord_user = c(6,1,3,4,2,5),
        # input user specified lower level class orders
        reord_user_high = c(1,4,3,2),
        # input user specified higher level class orders
        extout = TRUE)
        # print extensive output to see full model details
```

Software demonstration: multilevLCA R package Nonparametric approach with covariates

```
test_mod_consequence3<-multiLCA(
        data = dt_new,  #specify the dataset
        Y = c("con_1","con_2","con_3", "con_4","con_5",
                      "con_6","emb_rude","con_9",
                      "agg_hurt","hungover"),
        #specify the indicators
        iT = 6,
        # lower level number of classes
        id_high = "ID_temp",
        # higher level grouping variable
        iM = 4,
        # higher level number of classes
        Z = "pos_affect",
        #lower level covariate
        Zh = "underage",
        # higher level covariate
        NRmaxit = 10000,
        # Maximum number of iterations for
        # Newton-Raphson algorithm. Default: 100.
        NRtol = 1e-20,
        # Tolerance for Newton-Raphson algorithm.
        # Default: 1e-6.
        maxIter = 1e6,
        # Maximum number of iterations for
        # EM algorithm. Default: 1e3.
        tol = 1e-20,
        # Tolerance for EM algorithm.
        # Default: 1e-8.
        sequential = FALSE,
        numFreeCores = 8,
        extout = TRUE,
        # print extensive output to see full model details
        reord = TRUE,
        # customizing reference class
        reord_user = c(6,1,3,4,2,5),
        # input user specified lower level class orders
        reord_user_high = c(1,4,3,2))
        # input user specified higher level class orders
```

Software demonstration: multilevLCA R package
Try it with your data

- Important notes:
  - Indicators: consecutive integers from 0
  - Higher-level grouping IDs: integers
  - Avoid local maxima: customizing *maxIter*, *tol*, *NRmaxit*, and *Nrtol*

```
NRmaxit = 10000,
# Maximum number of iterations for
# Newton-Raphson algorithm. Default: 100.
NRtol = 1e-20,
# Tolerance for Newton-Raphson algorithm.
# Default: 1e-6.
maxIter = 1e6,
# Maximum number of iterations for
# EM algorithm. Default: 1e3.
tol = 1e-20,
# Tolerance for EM algorithm.
# Default: 1e-8.
```

```
unique_random_IDs <- sort(unique(dt_new$randomID))
dt_new$ID_temp <- NA
for (i in 1:length(unique_random_IDs)) {
  dt_new$ID_temp[which(dt_new$randomID==unique_random_IDs[i])] <- i
}
```

Software demonstration: multilevLCA R package
Try it with your data

- Helpful features:
  - Changing reference classes: *reord* statements
  - Extensive output: *extout* = *TRUE* for model details

```
> View(test_mod_consequence1)
```

  - *Help(multiLCA)* : always a good, handy reference!

```
reord = TRUE,
# customizing reference class
reord_user = c(6,1,3,4,2,5),
# input user specified lower level class orders
reord_user_high = c(1,4,3,2))
# input user specified higher level class orders
```

```
# print out lower level item response probabilities
level1 = data.frame(test_mod_consequence1[["mPhi"]])
#write.csv(level1, file =
"l1_approach5_nocov_4L2_20260416.csv")
level2 = data.frame(test_mod_consequence1[["mPi"]])
#write.csv(level2, file =
"l2_approach5_nocov_4L2_20260416.csv")
```

Time for questions and discussion

THANK YOU!

Stephanie T. Lanza, Ph.D.
slanza@psu.edu

Bethany C. Bray, Ph.D.
bcbray@uic.edu

Yuqi Shen, M.H.S.
yms5393@psu.edu