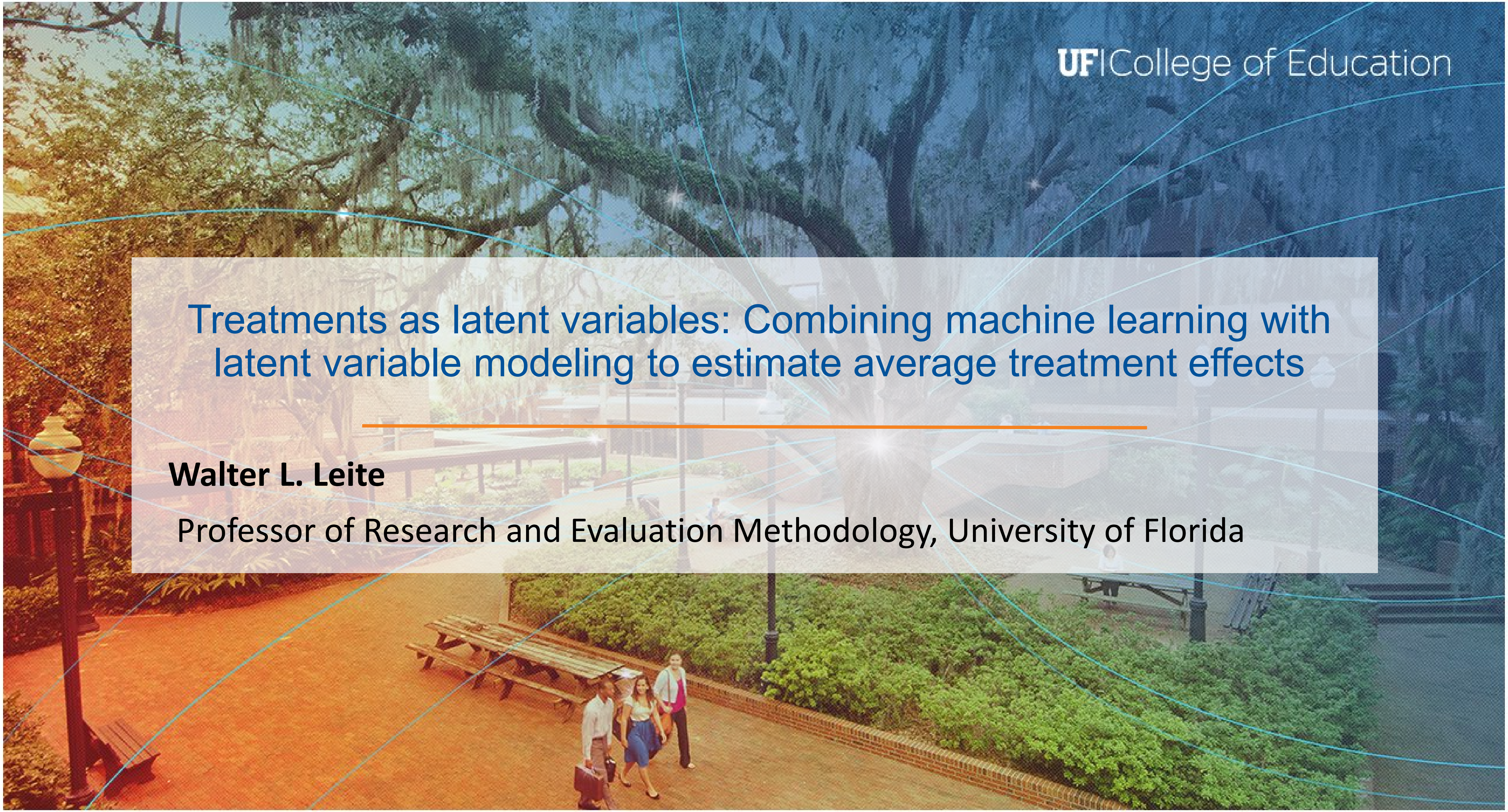


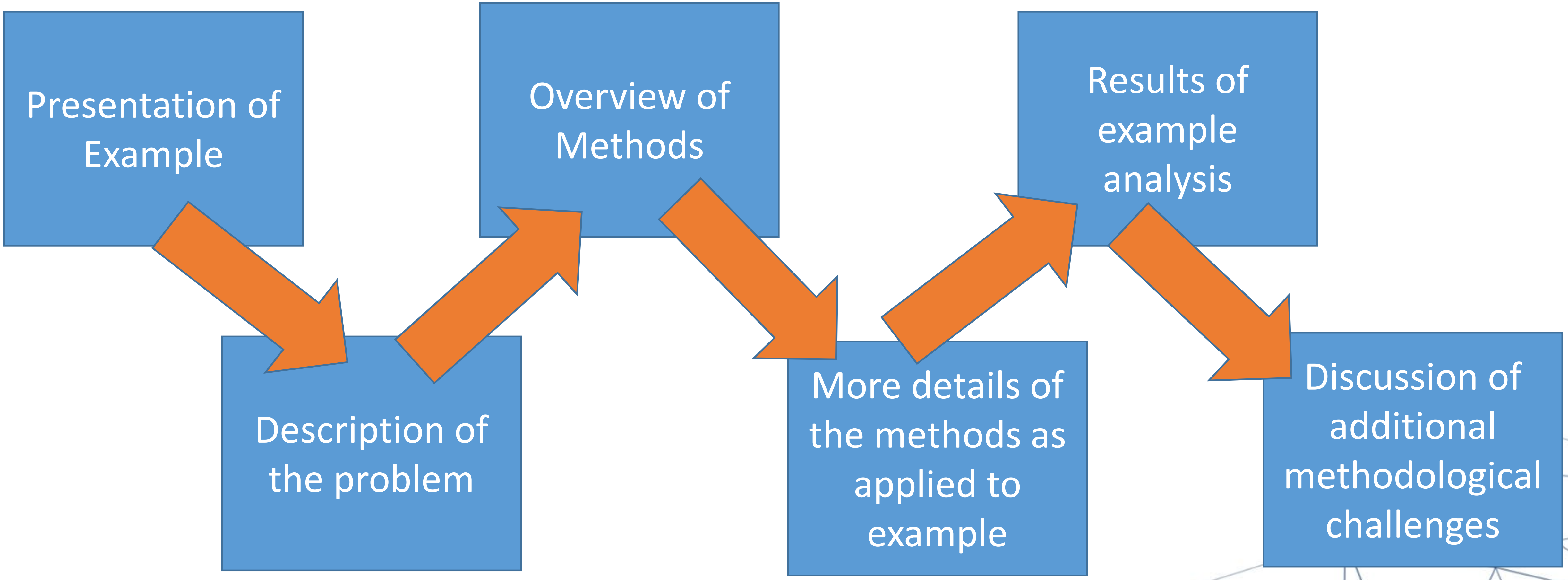
UF | College of Education

Treatments as latent variables: Combining machine learning with latent variable modeling to estimate average treatment effects

Walter L. Leite

Professor of Research and Evaluation Methodology, University of Florida

Structure of the presentation



Presentation of Example

Description of the problem

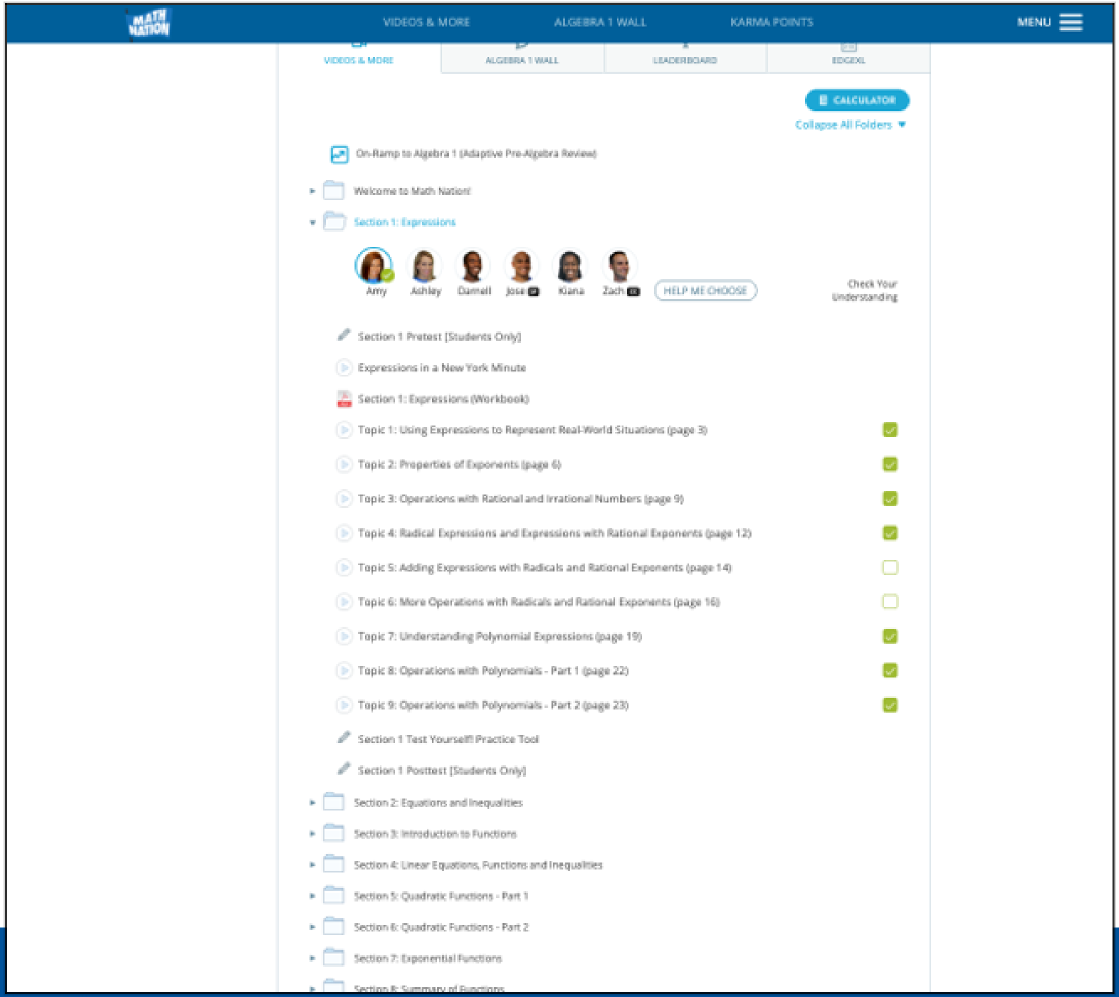
Overview of Methods

More details of the methods as applied to example

Results of example analysis

Discussion of additional methodological challenges

Example Virtual Learning Environment: Math Nation



On-Ramp to Algebra 1 (Adaptive Pre-Algebra Review)

Welcome to Math Nation!

Section 1: Expressions

- Section 1 Pretest [Students Only]
- Expressions in a New York Minute
- Section 1: Expressions (Workbook)
- Topic 1: Using Expressions to Represent Real-World Situations (page 3) ☑
- Topic 2: Properties of Exponents (page 6) ☑
- Topic 3: Operations with Rational and Irrational Numbers (page 9) ☑
- Topic 4: Radical Expressions and Expressions with Rational Exponents (page 12) ☑
- Topic 5: Adding Expressions with Radicals and Rational Exponents (page 14) □
- Topic 6: More Operations with Radicals and Rational Exponents (page 16) □
- Topic 7: Understanding Polynomial Expressions (page 19) ☑
- Topic 8: Operations with Polynomials - Part 1 (page 22) ☑
- Topic 9: Operations with Polynomials - Part 2 (page 23) ☑
- Section 1 Test Yourself! Practice Tool
- Section 1 Posttest [Students Only]

Section 2: Equations and Inequalities

Section 3: Introduction to Functions

Section 4: Linear Equations, Functions and Inequalities

Section 5: Quadratic Functions - Part 1

Section 6: Quadratic Functions - Part 2

Section 7: Exponential Functions

Section 8: Summary of Functions

- Math Nation is used by 950,000 students per year and over 30,000 teachers.
- It is aligned with the Florida Mathematics standards and is the adopted curriculum in several school districts.
- Its major components are practice problems, tutorial videos, and a discussion board.

The Problem: Quasi-experimental Evaluation of Virtual Learning Environments

- Virtual Learning Environments are very prevalent, but are difficult to evaluate rigorously:

  A. There is no control group

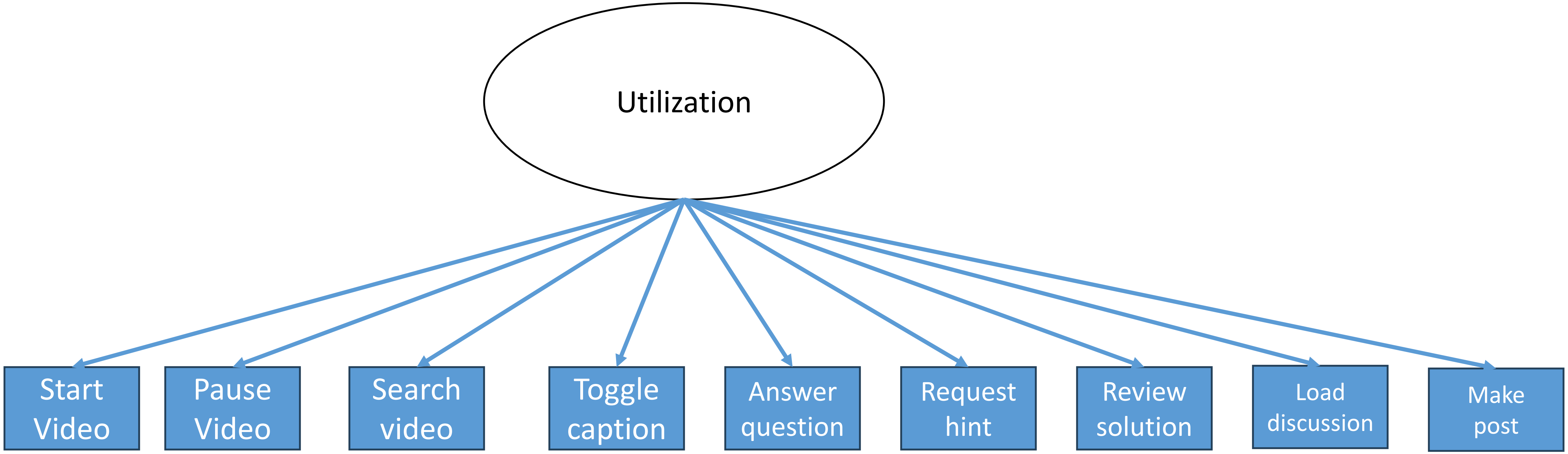
  B. There is large heterogeneity of usage

  C. Usage patterns are unknown

  D. Usage may be driven by students and/or teachers

Treatment as a latent variable

- The logs of student actions in the virtual learning environment (VLE) can be used as indicators of a latent variable that characterizes utilization.



Continuous or categorical latent variable

- Both continuous and categorical latent variables are not directly observed but have indicators that measure the latent variable with error.

- Models for continuous latent variables include exploratory factor analysis, confirmatory factor analysis, and item response theory models.

- Models for categorical latent variables include latent class models, latent profile models, cognitive diagnosis models, and Bayesian knowledge tracing models.

Latent Classes in Learning Analytics Research

- Latent classes can be used to summarize to classify VLE users into mutually-exclusive groups based on system logs.
- Latent class analysis (LCA) estimates the probabilities of subjects being classified into certain latent classes.
- LCA allows for uncertainty of class membership to be evaluated.
- LCA can include covariates that predict latent classes, as well as distal outcomes of latent classes.

Propensity Score Methods for Latent Classes

- Rubin’s causal model can be used to define potential outcomes of membership in each latent class.

- We can estimate the average treatment effect of all study participants being in one class versus in another class under the assumption of weak unconfoundedness (Imbens, 2000)

- To remove selection bias, propensity score methods can be used to balance covariate distributions across latent classes.

The Latent Class Analysis Model

Probability of a response to item i by subject s

$$\pi_s^i = \sum_{c=1}^{C} \pi_{sc}^{C} \pi_{ic}^{Ti|C}$$

Conditional response probabilities

Latent class probability

How to compare models

- Akaike Information Criterion

$$AIC = -2\log L + 2p$$

- Bayesian Information Criterion

$$BIC = -2\log L + p\log(n)$$

- Consistent Akaike Information Criterion

$$CAIC = -2\log L + p[\log(n) + 1]$$

- Sample-adjusted Bayesian Information Criterion

$$aBIC = -2\log L + p\log[(n+2)/24]$$

- Lo-Mendell-Rubin likelihood ratio test (LMR test, Tech11)
- Bootstrap Likelihood Ratio Test

Research Questions

- What are latent classes that summarize the student and classroom usage patterns of Math Nation?

- What is the average treatment effect of students belonging to a Math Nation usage latent class on the Algebra EOC assessment scores?

- Is the average treatment effect of students belonging to a Math Nation usage latent class moderated by the usage latent class of their classroom?

Sample

- The sample consisted of 42,698 students and 1,020 teachers from 631 schools in Florida.

- System log data for the Spring 2017 was obtained from Math Nation;

- Variables obtained from the Florida Department of Education included grade levels, race, free or reduced lunch status, exam status, previous Florida Standard Assessment (FSA) score (2015~2016 academic year), and Algebra I EOC scores (2016~2017 academic year). .

Steps of Analysis

1) Latent Class Analysis: Identify latent classes at student and classroom levels according to VLE usage indicators based on log data;

2) Propensity Scored Analysis: Estimate inverse probability weighs and evaluate covariate balance.

3) Multilevel mixture model: Estimate the **average treatment effect** of class memberships of students moderated by the latent class membership of their classrooms.

- Posterior probabilities of class membership: Adjust for uncertainty of student and classroom latent classes;

- Inverse probability weights: Adjust for non-random selection into classes due to observed student, teacher and classroom covariates;

- Cluster robust standard errors: Adjust for clustering effects on the outcome;

Latent Class Analysis

1) Set model selection criterion: Integrated complete data likelihood implemented in the *VarSelLCM* package (Marbac, M. and Sedki, M., 2017a; 2017b) of R.

$$ICL(m, K) = log\boldsymbol{f}\left(\boldsymbol{x}, \hat{\boldsymbol{z}} \middle| m, K, \hat{\theta}\right) - \frac{\upsilon_{m,K}}{2} logn$$

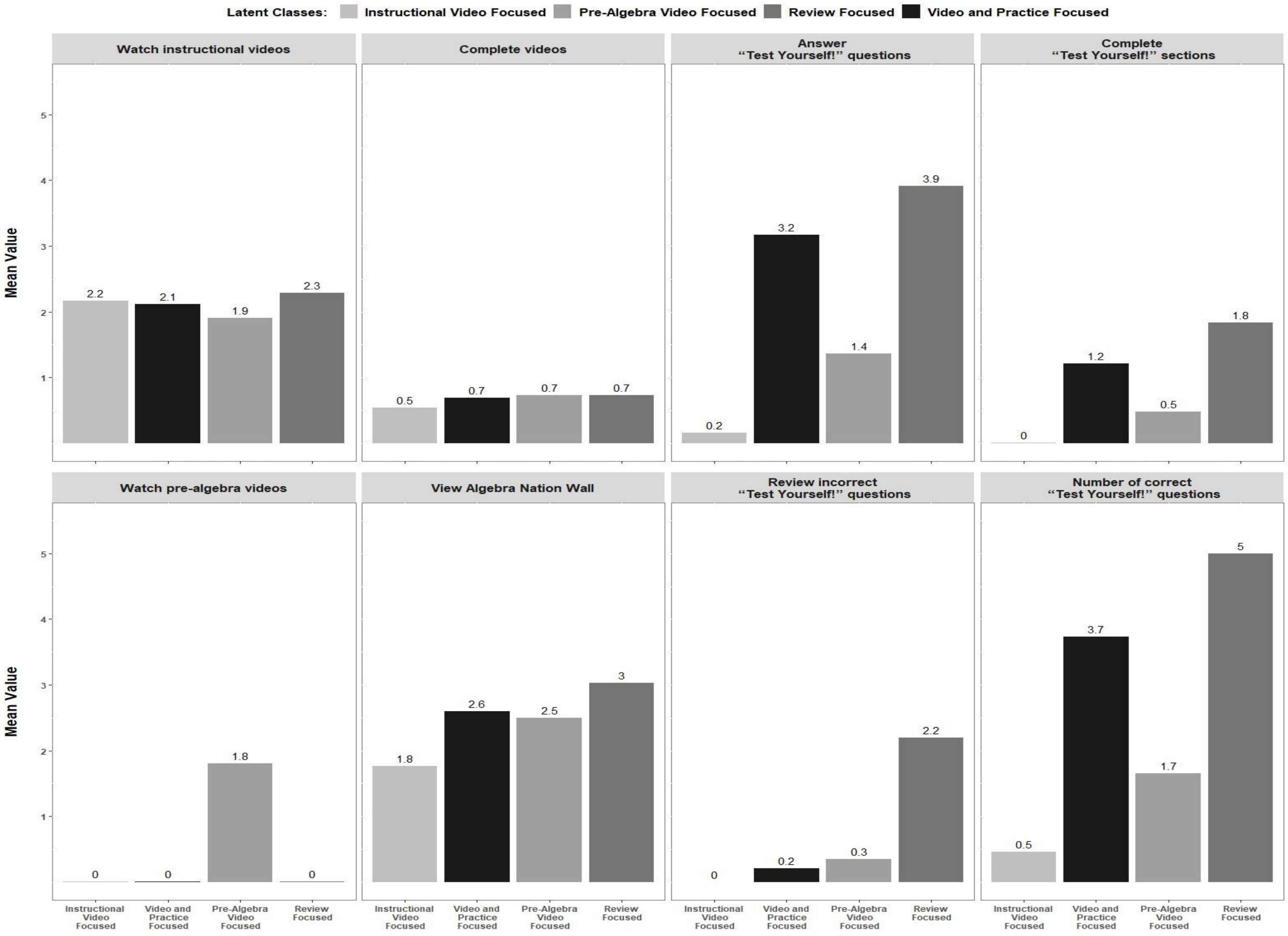
2) Determine the indicators and number of classes: hybrid Ant Colony Optimization (hACO) algorithm (Jing, Kuang, Leite & Huggins-Manley, 2019)

3) Calculate posterior probabilities of class membership and most likely class membership.

4) Interpret classes based on indicator probabilities.

Results of LCA for students

4 classes with 8 out of 11 indicators selected



Propensity scores Analysis

- The propensity score is defined as a conditional probability of treatment assignment, given observed covariates (Rosembaum & Rubin, 1983);

$$e(x_i) = P(Z_i = 1 \mid \mathbf{X})$$

- If the propensity score was correctly specified, balancing the treatment and control groups with respect to propensity score also balances them with respect to distributions of covariates;

Advantages of Propensity Score Methods over Conditioning on Covariates

- Smaller models where fewer parameters are estimated;

- Linearity assumptions are not made;

- Problem of differences in distributions of covariates for treatment and control groups is eliminated.

Strong Ignorability of Treatment Assignment

- For binary treatments: The treatment assignment is independent of the potential outcome distributions of all treatment versions, given observed covariates.

$$\left[Y_i^1, Y_i^0 \right] \perp Z_i \mid \mathrm{X}_i$$

- It also requires that for every value of Z, the probability of treatment assignment is neither zero nor one

$$0 < p(\mathrm{Z}_i = 1 \mid \mathrm{X}_i) < 1$$

Weak Unconfoundedness
(Weak Ignorability of Treatment Assignment)

- For multiple treatment versions: Each treatment is independent of its own potential outcome distribution, given observed covariates.

$$Z_i^j \perp Y_i^j \mid X$$

- It also requires that for every value of Z, the probability of assignment to each treatment is neither zero nor one

$$0 < p(\mathrm{Z}_i = j \mid \mathrm{X}) < 1$$

Generalized Propensity Score (GPS)

- It is the conditional probability of receiving a particular level of treatment given covariates.

$$P(\mathrm{Z}_i = j \mid X)$$

- Weak ignorability assumption with GPS:

$$Z_i^j \perp Y_i^j \mid P(\mathrm{Z}_i = j \mid X)$$

Steps of Propensity Score Analysis

Step	Objective
Data preparation	Obtain complete data that is ready for analysis
Propensity score estimation	Obtain propensity scores for treated and untreated individuals
Propensity score method implementation	Implement a strategy to balance treated and untreated covariate distributions using propensity scores
Covariate balance evaluation	Determine the degree to which balance of covariate distributions between treated and untreated was achieved
Treatment effect estimation	Estimate the treatment effect and its standard error
Sensitivity analysis	Determine how strong the effect of an omitted covariate would have to be for the significance test of the treatment effect to change

Estimation of Propensity Scores

- Different methods for estimating PS can be used:
  - *Statistical models*: logistic regression, probit regression
  - *Machine learning algorithms:*
    classification trees, boosting, bagging, random forests, support vector machines, neural networks, super learner.

Estimation of Generalized Propensity Score with Data Mining

- Create a dummy indicator for each treatment version

- Predict each dummy indicator

- The generalized propensity score for each individual is the predicted probability of the treatment version that the individual was exposed to.

Generalized Boosted Modeling

- Boosting is a general method to improve a predictor by reducing prediction error.
- GBM for propensity score estimation improves prediction of the logit of treatment assignment:

$$\text{logit}(Z_i = 1 \mid X)$$

- Starting value: $\log[\bar{Z} / (1 - \bar{Z})]$
- Regression trees are used to minimize the within-node sum of squared residual:.

$$Z_i - e_i(X)$$

Stopping GBM

- There is no defined stopping criterion, so errors decline up to a point and then increase,

- For propensity score estimation, McCaffrey et al. (2013; 2004) recommended using a measure of covariate balance, to stop the GBM algorithm the first time that a minimum covariate balance is achieved.

- There is no guarantee that better covariate balance would not be achieved if the algorithm runs additional iterations.

Generalized Propensity Score for a Latent Class

The generalized propensity score of a latent class is defined as the sum of fitted probabilities weighted by the posterior probabilities for an individual appearing in each latent class (Bray, Dziak, Patrick, & Lanza, 2019).

$$\widehat{\pi}_i = \sum_{t=1}^{T} P(T = t|\boldsymbol{X}_i)P(T = t|\, u_i)\,,$$

We estimated the fitted probabilities via Generalized Boosted Modeling (GBM), using the *twang* package (Ridgeway, McCaffrey, Morral, Griffin, & Burgette, 2017).

We used 11 covariates to predict student latent classes and 5 covariates to predict classroom latent classes.

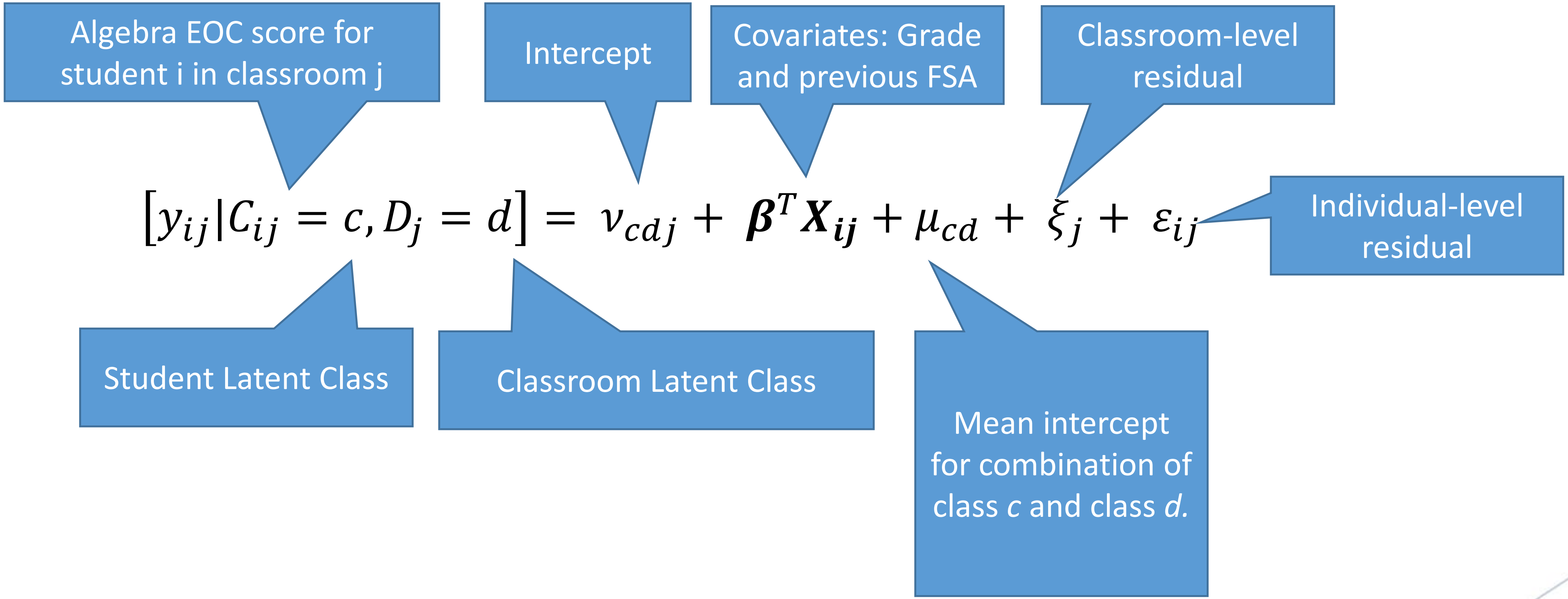
Inverse Probability of Treatment Weighting with the generalized propensity score

$$w_i = \frac{1}{\hat{\pi}_i}$$

Covariate balance was evaluated between all possible pairs of latent classes using absolute standardized mean differences.

Only two student-level covariates exceeded the 0.25 cutoff (0.28 for grade level and the 0.27 for previous FSA score)

Estimation of Treatment Effect: Multilevel Mixture Model



Algebra EOC score for student i in classroom j

Intercept

Covariates: Grade and previous FSA

Classroom-level residual

$$\left[y_{ij}|C_{ij}=c,D_j=d\right]= \nu_{cdj}+\boldsymbol{\beta}^T\boldsymbol{X}_{\boldsymbol{ij}}+\mu_{cd}+\xi_j+\varepsilon_{ij}$$

Individual-level residual

Student Latent Class

Classroom Latent Class

Mean intercept for combination of class *c* and class *d.*

The model was estimated with robust maximum likelihood estimation with two vectors of inverse probability weights (students and classroom weights) and cluster-robust standard errors

Average Treatment Affects

The mean intercept μ_{cd} is an estimate of $E[Y(c,d)]$, which is expected potential outcome of a student belonging to student class *c* and classroom class *d*. We estimated:

$$ATE_{c1d} = E[Y(C=c, D=d)] - E[Y(C=1, D=d)]$$

ATE_{c1d} is the average treatment effect of students participating in student latent class *c* instead of student latent class 1, which is the reference student latent class, for classroom latent class *d*.

The interaction effect is the difference in ATE_{c1d} between a classroom latent class *d* and classroom latent class 1.

$$I_{cd1} = E[Y(C=c, D=d)] - E[Y(C=1, D=d)] - \ E[Y(C=c, D=1)] - E[Y(C=1, D=1)]$$

Accounting for uncertainty of class membership

We used Vermunt´s (2010) three-step method:

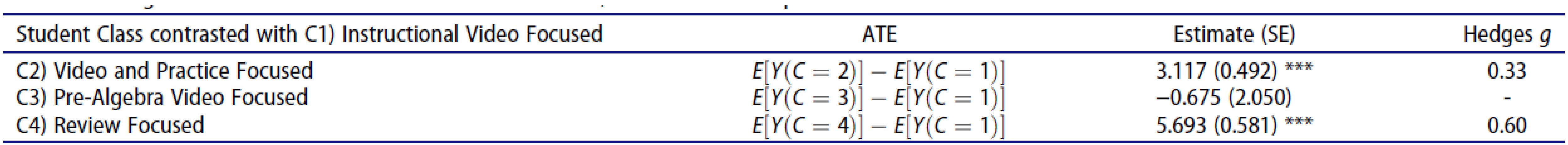
1. From the unconditional LCA, two vectors containing the most likely class membership of students and classrooms were saved.

2. The logits of the posterior probabilities of latent classes were collected

3. The multilevel mixture model was estimated with the two vectors of most likely class membership for student and teachers used as nominal class indicators, and with mean logits of the latent classes fixed to the logits of the posterior probabilities collected in step two.

Average Treatment Effect Estimates

Student Latent Classes



Student Class contrasted with C1) Instructional Video Focused	ATE	Estimate (SE)	Hedges *g*
C2) Video and Practice Focused	$E[Y(C=2)]-E[Y(C=1)]$	3.117 (0.492) ***	0.33
C3) Pre-Algebra Video Focused	$E[Y(C=3)]-E[Y(C=1)]$	−0.675 (2.050)	-
C4) Review Focused	$E[Y(C=4)]-E[Y(C=1)]$	5.693 (0.581) ***	0.60

Conclusions: Challenges of Causal Evaluation using system log data

1. How to address measurement error and outliers
2. How to select indicators of latent classes while simultaneously performing class enumeration.
3. Construct validity of the interpretations of the latent classes.
4. How to account for selection bias in VLE usage due to observed and unobserved confounders.
5. How to account for the uncertainty of class membership

Thank you!

walter.leite@coe.ufl.edu

https://virtuallearninglab.org/

https://www.practicalpropensityscore.com/

UF | College of Education