UF | College of Education

Heterogeneity of treatment effects of a video recommendation system for algebra

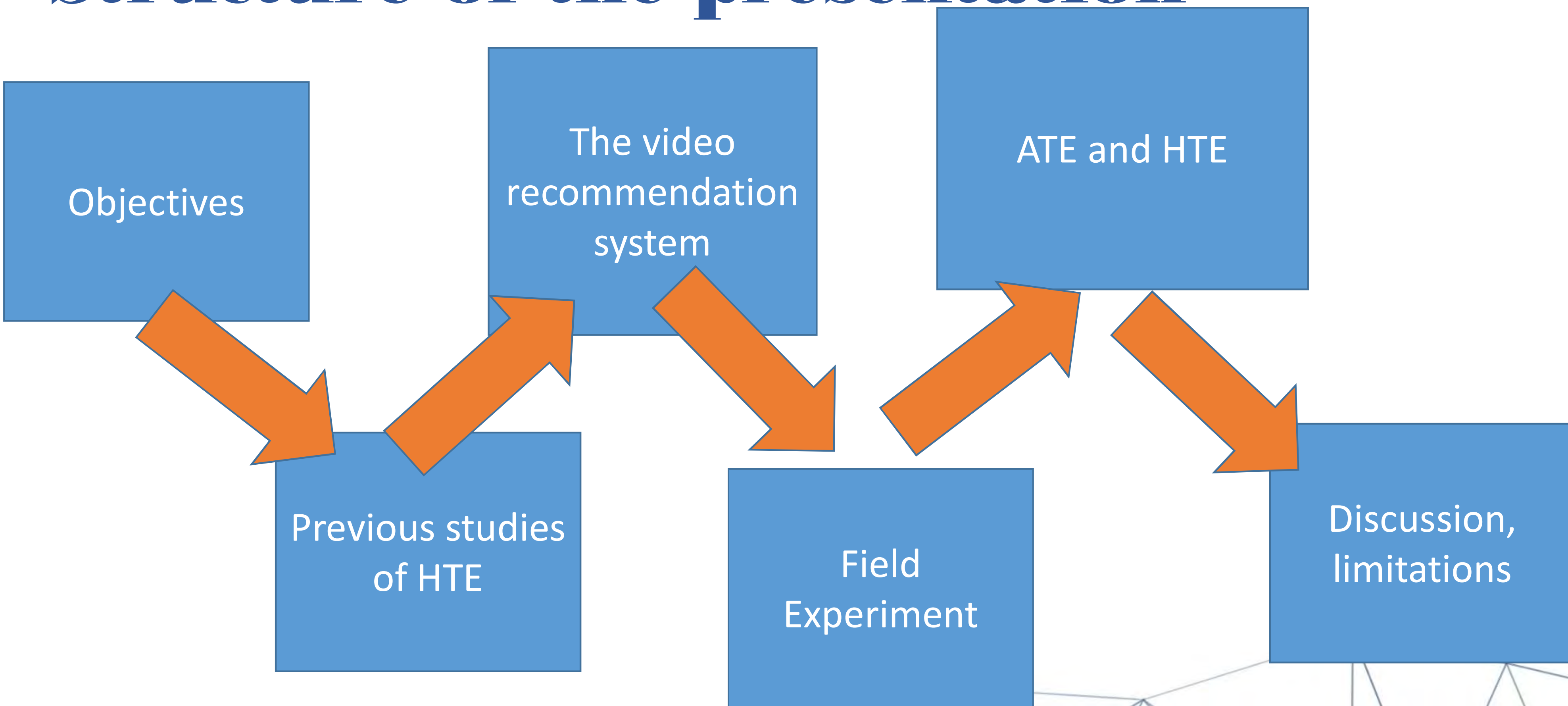
Presenter: Walter L. Leite, Director of the Virtual Learning Lab

Authors: Walter L. Leite, Huan Kuang, Zuchao Shen, Nilanjana Chakraborty, George Michailidis, Sidney D’Mello & Wanli Xing

Acknowledgements:

- The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C160004 to the University of Florida. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Structure of the presentation



Objectives and research questions

- <u>Objective:</u> Extend a previous evaluation (Leite at al., 2022) of a novel video recommendation system for an online Algebra learning platform, Algebra Nation by examining heterogeneity of treatment effects (HTE).

- <u>Research questions:</u>
  1) What are the effects of the recommendation system on learning outcomes both within the platform and from standardized tests compared to a control group;
  2) Is there substantial HTE of the video recommendation system?
  3) What student characteristics predict the HTE?

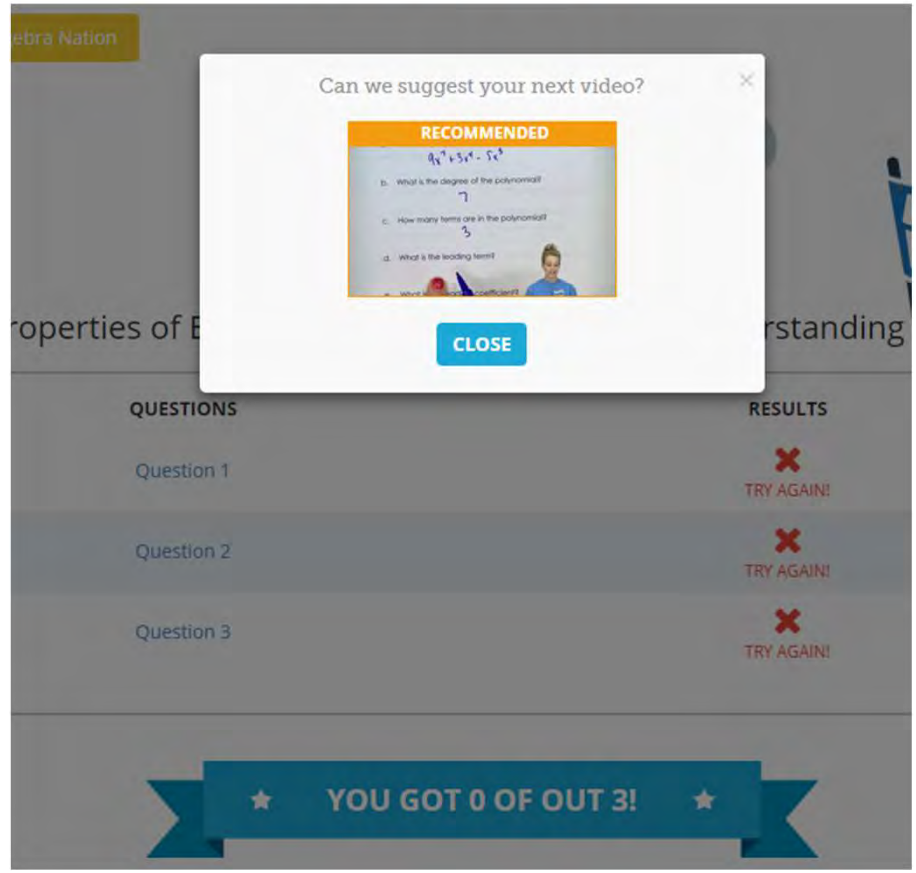
Previous work on HTE for ITS

- Examination of HTE in intelligent tutoring systems (ITS) has been performed mostly through meta-analysis, but only two out of six meta-analysis considered HTE due to student characteristics:
  - Ma et al. (2014) found that ITS used with middle-school and post-secondary students had higher effects than with elementary and high-school students.
  - Steenbergen-Hu and Cooper (2013) found that the effectiveness of ITS for helping students drawn from the general population was greater than for helping low achievers.

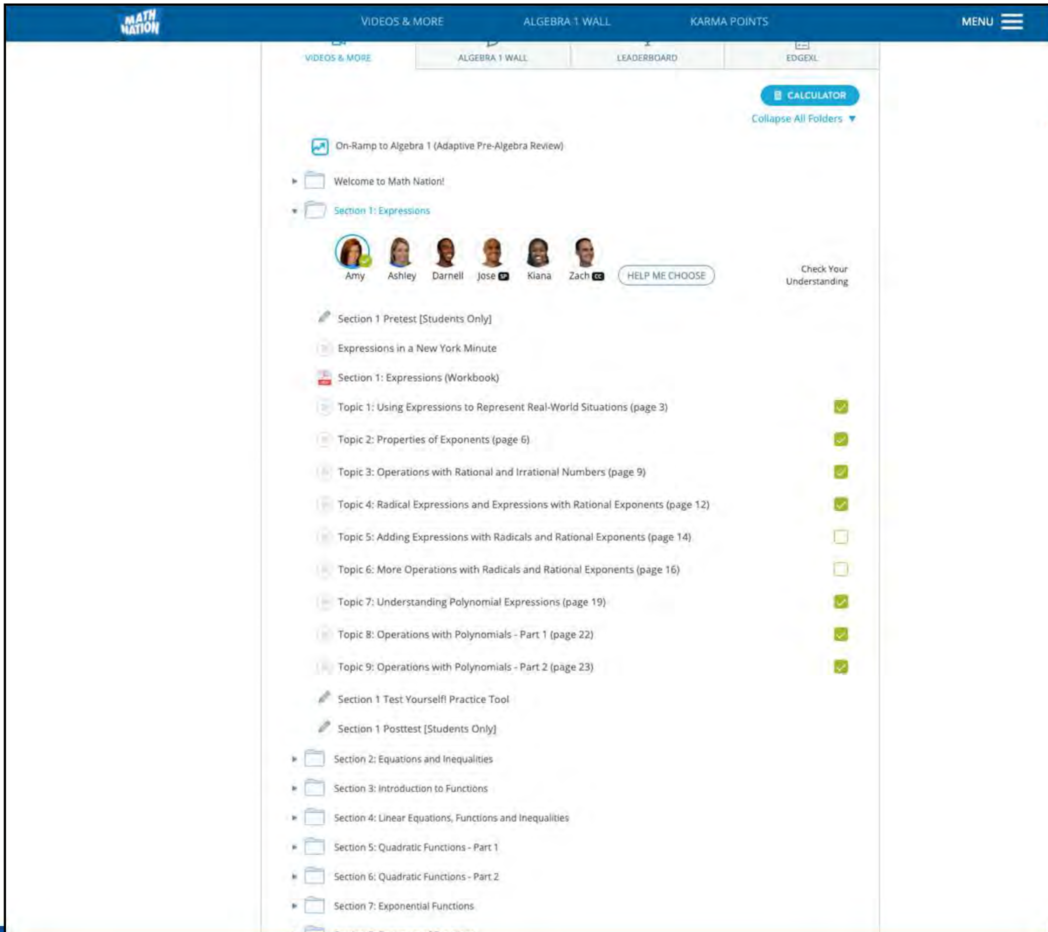
Current Study: The Video Recommendation System

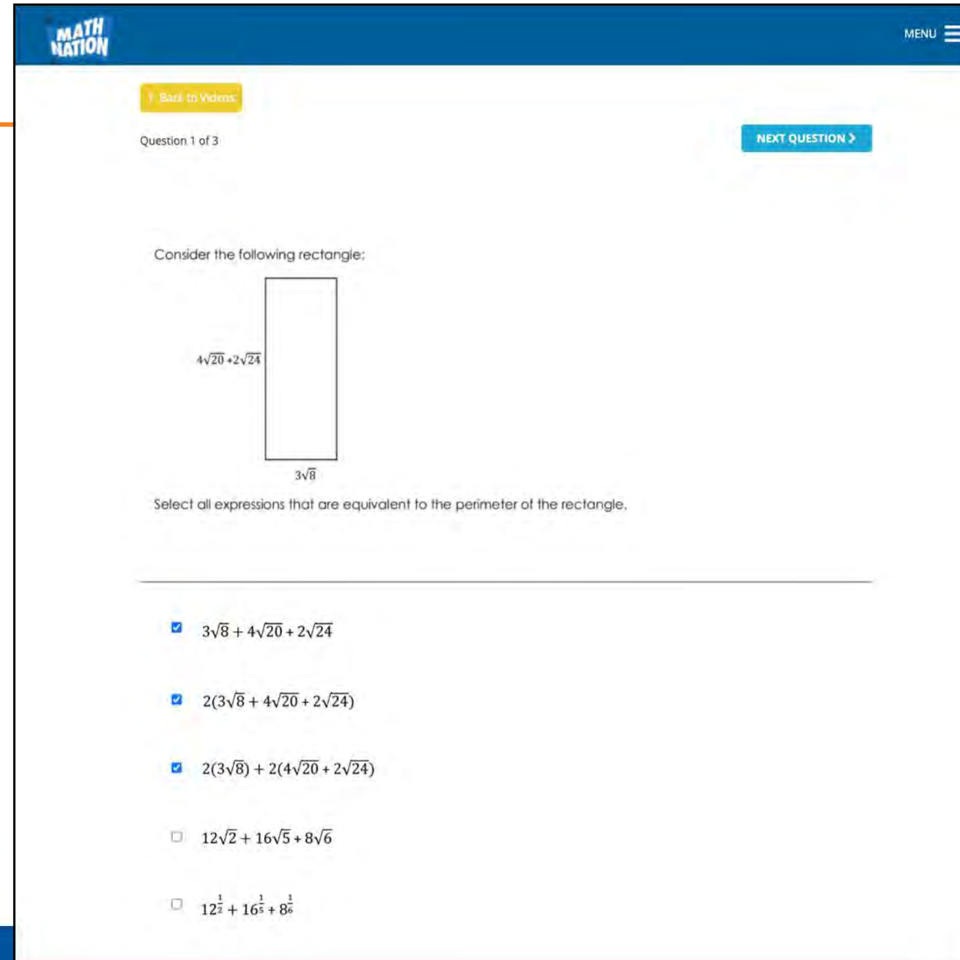
<u>Objective:</u> Provide students with a personalized video recommendation that takes into account their current **knowledge** as well as their **engagement** with the system.

Video Recommendation Screen in Math Nation

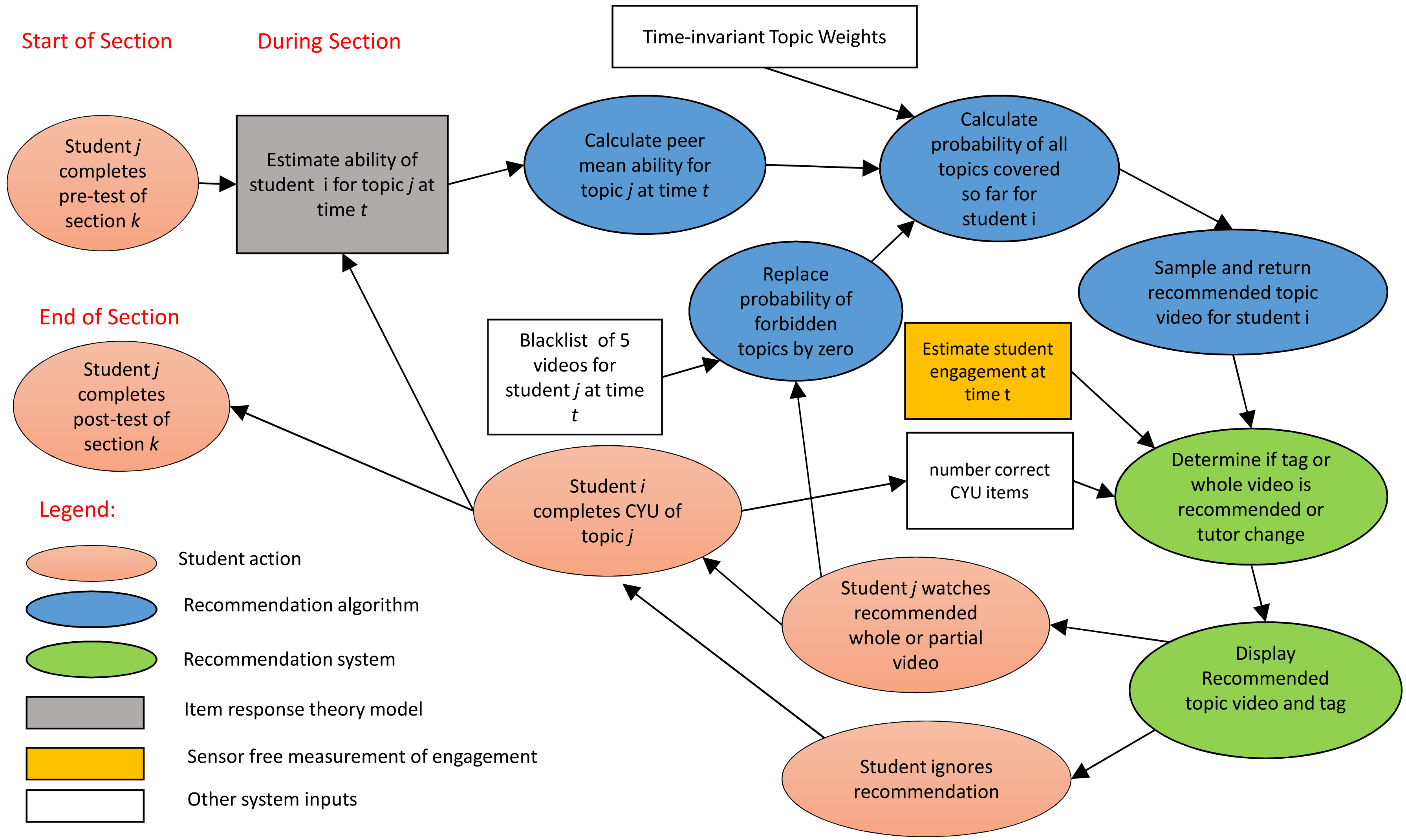


Context: Math Nation





The video recommendation flowchart



The algorithm for new videos

ALGORITHM 1. New Video Recommendation Policy for Student *i*

Inputs: *initial ability estimates* $\{a_{ij}(0)\}$, $1 \leq i \leq n, 1 \leq j \leq r$.

Output: *sequence of recommended videos* $\hat{j}\,(t) \in \{1, \ldots, r\}, t \geq 0$

for $\boldsymbol{t = 0, 1, \ldots}$ **do**

Compute peer ability-estimates

$$b_j(t) = n^{-1} \sum_{i=1}^{n} a_{ij}(t).$$

Compute the probability distribution $\{p_j(t)\}, j = 1,2, \ldots, r,$

$$p_j(t) = \frac{exp\,[-w_j\,(a_{ij}(t) - b_j(t))]}{\sum_{j=1}^{r} exp\,[-w_j\,(a_{ij}(t) - b_j(t))]}.$$

Sample $\hat{j}\,(t)$ *from the distribution* $\{p_j(t)\}_{,1 \leq j \leq r}$.

Read $\{a_{ij}(t+1)\}$ $,1 \leq i \leq n$ $,1 \leq j \leq r$ *from the database.*

end for

The video recommendation system

Three categories of video recommendation:

C1: view new video as determined by the recommendation algorithm;

C2: review segment of current video that is most related to the questions that the student answered incorrectly (by expert review) from a new tutor;

C3: view next video in curriculum sequence.

CYU quiz score (0 to 3)	Engagement score (0 <3 low; >= 3 high)	Probability of recommendation Category		
		C1	C2	C3
0	low	0.9	0.1	0.0
0	high	0.7	0.3	0.0
1	low	0.7	0.3	0.0
1	high	0.5	0.5	0.0
2	low	0.5	0.5	0.0
2	high	0.3	0.7	0.0
3	low	0.3	0.0	0.7
3	high	0.1	0.0	0.9

Field Experiment

- Sample: 2,995 middle and high-school students from 54 teachers in 42 schools. 35.1% of students were attending school campuses in person.
- Students were enrolled in Algebra 1 or Algebra 1 Honors courses in the 2020-2021 academic year.
- The study lasted from January to June 2021
- Students within teachers were randomly assigned to recommendation system or control (always view next video in sequence). Treatment assignment was blind to students and teachers.
- Two outcomes: 10-question post-tests, high-stakes Algebra 1 End-of-Course assessment

Average Treatment Effects (Intention to Treat)

The ATE was estimated with a multilevel model:

Student Level: $y_{ij} = \beta_{0j} + \delta_j T_{ij} + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma^2_1)$

Teacher Level: $\beta_{0j} = \gamma_{00} + u_{0j}$

$\delta_{1j} = \delta + u_{1j}$

	Post-test			Algebra 1 EOC assessment		
	ICC	Hedges G	p-value	ICC	Hedges G	p-value
	0.341	0.330	< 0.001	0.582	0.170	< 0.001

Heterogeneity of Treatment Effects

- Two-stage cluster-robust causal forest (Wager & Athey, 2018) to measure the importance of each predictor and to estimate the individual conditional average treatment effect (iCATE):
  1. Entire data set *S* was randomly divided into two subsamples
  2. In the first stage, a pilot causal forest was trained on all features with first subsample. Variable importance was computed as a depth-weighted average of the number of splits on the variable of interest.
  3. In the second stage, another forest was trained on the second subsample to estimate iCATEs. We only selected for the second stage those variables whose importance exceeded the median of the variable importance.

Variables with Importance Above Median (from 17 variables)

Variables	Average Variable Importance (Post-test)	Average Variable Importance (EOC)
pre-test ability	0.241	0.241
Followed rate	0.228	0.229
Absent days	0.116	0.115
Ethnicity (Hispanic)	0.081	0.081
Percent distance learning	0.075	0.079
Free or reduced-price lunch	0.068	0.068
Sex	0.045	0.045

Modeling relationship between iCATEs and important variables

- We used the following multilevel model:

$$y_{ij} = \beta_{0j} + \sum_{k=1}^{d} \delta_{jk} X_{ijk} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(\mathbf{0}, \boldsymbol{\sigma}^2)$$

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\delta_{1j} = \delta + u_{1j}$$

Variables	Post-test	EOC score
Pre-test ability	0.077*** (0.013)	0.071*** (0.009)
Followed rate	0.152*** (0.038)	0.455*** (0.027)
Sex indicator	-0.006 (0.013)	-0.033 (0.009)
Absent days	-0.003 (0.001)	0.002** (0.001)
Free or reduced-price lunch indicator	0.144 *** (0.014)	-0.004 (0.011)
Percent distance learning	0.001 *** (0.001)	0.001*** (0.001)
Ethnicity indicator (Hispanic)	0.040 ** (0.015)	0.107*** (0.010)

Discussion

- The HTE due to pre-test ability could be because students with higher previous achievement having better SRL skills;

- The HTE due to free-and-reduced lunch eligibility of the students could indicate that economically disadvantaged students benefited more from the recommendation system

- Hispanic students had higher iCATEs on both post-test and the EOC than non-Hispanic students, which may be associated with cross-cultural differences, differences in teacher orchestration of technology in classrooms, or school level contextual differences.

Limitations

- The current study does not clarify the specific mechanisms by which some subgroups of students (e.g. Hispanic students and free-reduced lunch eligible students) benefited more from the recommendation system.
- The current study did not include teacher variables in the prediction of HTE, such as survey variables indicating when and how teachers used the VLE with their students, or school contextual variables such as percent of minority students, expenditures per pupil, and percent of students in poverty.
- A previous competition of machine learning-based HTE detection methods (Carvalho et al., 2019) has shown that results can vary substantially across methods, but the current study used a single method.

References

- Carvalho, C., Feller, A., Murray, J., Woody, S., & Yeager, D. (2019). Assessing Treatment Effect Variation in Observational Studies: Results from a Data Challenge. *Observational Studies*, *5*(2), 21-35. https://doi.org/10.1353/obs.2019.0000

- Leite, W.L., et al. A novel video recommendation system for algebra: An effectiveness evaluation study. in LAK22: 12th International Learning Analytics and Knowledge Conference (LAK22). 2022. Online, USA: ACM, New York, NY, USA. https://dl.acm.org/doi/10.1145/3506860.3506906

- Wager, S., & Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, *113*(523), 1228-1242. https://doi.org/10.1080/01621459.2017.1319839

Thank you!

More information and contact:

https://virtuallearninglab.org/