CausalEDM: Linking Innovations in Instructional Design and the Complex Behaviors that Underlie Learning Processes and Outcomes

Kirk P. Vanacore
University of Pennsylvania
vanacor1@upenn.edu

Anthony F. Botelho
University of Florida
a.botelho@ufl.edu

Avery H. Closser
University of Florida
avery.closser@ufl.edu

Adam C. Sales
Worcester Polytechnic Institute
asales@wpi.edu

Neil T. Heffernan
Worcester Polytechnic Institute
nth@wpi.edu

ABSTRACT

Causal inference is becoming a critical focus within the field of Educational Data Mining (EDM), providing insights into how instructional innovations influence learning behaviors and outcomes. The CausalEDM workshop at EDM 2025 aims to synthesize past research, showcase ongoing work, and explore future directions in causal inference methodologies. This workshop will feature a retrospective on causal research in EDM, highlighting advancements in experimental design, effect heterogeneity, and quasi-experimental approaches. Discussions will include the methods for estimating the impact of Generative AI on student engagement, the integration of machine learning in causal modeling, and emerging methodological challenges. The event will include expert panels, paper presentations, and lightning talks, fostering collaboration between causal inference specialists and the broader EDM community. Through these discussions, the workshop seeks to explore causal methodologies, expand their applications, and address key research gaps in understanding the interplay between instructional design and learning processes.

Keywords

Causal Inference, Treatment Effects, Causal Modeling, Data Mining, Experimental Design

1. INTRODUCTION

Since the inception of CausalEDM five years ago, educational data mining (EDM) has made substantial progress in causal methodologies and applications. The proliferation of A/B testing and experiments can be seen in various works across digital learning platforms [13, 9, 18]. Some platforms have established systems for researchers to conduct *en vivo* experiments (e.g., E-trails[1]; UpGrade[2]). There have also been various methodological advancements made by my EDM researchers (e.g., Fully Latent Principal Stratification [19, 16, 14], remanent Leave-One-Out Potential outcomes [12, 15], accounting for nested data structure [3], and demonstrating how to supplement quantitative methods with EDM techniques to observe causal relations within experimental datasets [2]). EDM researchers have also identified effect heterogeneity in digital learning platforms [8, 1]. Furthermore, the EDM community has begun to explore causal estimating quasi-experimental or non-experimental data [7]. This year, we will provide a cohesive retrospective on prior causal work in EDM, a sampling of current projects, and a discussion of future directions.

We see a potential area of growth in the development and applications of causal methods that allow us to understand how innovations in institutional design affect students learning behaviors and epistemic effects, which then impact these innovations' efficacy. One example is how Generative Artificial Intelligence (GenAI) impacts students' engagement with educational materials. We have seen that GenAI can produce high-quality institutional content [4, 5] that can positively impact learning [11, 10]. We have also seen that GenAI can hinder students' metacognition and self-regulated learning [6, 17]. For these innovations to be truly impactful, we must look beyond binary classifications of effectiveness and do more to understand the causal paths that link implementation and engagement with the innovations to their efficacy.

This workshop will include a presentation synthesizing prior causal work in EDM and positing some areas for growth. This will lead into a panel discussion on future directions for causal inference research. The panel will include experts in causal inference active in the EDM field as well as prominent EDM researchers with focuses across various areas (e.g., affect and behavior detection, GenAI applications, etc.). Next, we will have presentations on papers of more developed work and lighting talks featuring emerging ideas and in-progress works. A discussion of questions and feedback for the presenters will follow each presentation. We will end with an open, moderated discussion to identify current

[1] https://etrials.assistments.org/
[2] https://www.upgradeplatform.org/

Kirk Vanacore, Anthony F. Botelho, Avery Closer, Adam Sales, and Neil Heffernan. CausalEDM: Linking Innovations in Instructional Design and the Complex Behaviors that Underlie Learning Processes and Outcomes. In Caitlin Mills, Giora Alexandron, Davide Taibi, Giosuè Lo Bosco, and Luc Paquette (eds.) Proceedings of the 18th International Conference on Educational Data Mining, Palermo, Italy, July, 2025, pp. 705–708. International Educational Data Mining Society (2025).

https://doi.org/10.5281/zenodo.15870314

Table 1: Tentative Schedule (Half-Day)

0:00-0:20	Introductions
0:20-1:00	CausalEDM: Retrospective and Future Directions
1:00-1:30	Panel-Mediated Discussion
1:30-2:30	Three Paper presentations (10 min presentations, 10 min discussions)
2:30-3:10	Four Lighting Talks (5 minutes per talk, 5 minute discussions)
3:10-3:20	Closing Discussion: Causal Problems and Opportunities for EDM

and future needs for causal research in EDM.

2. CONTENT AND THEMES

The workshop will cover three themes of causal inference in EDM: *Empirical Studies* include applications of causal inference in applied EDM-related settings. These studies can include but are not limited to A/B tests, randomized control trials, and observational studies. *Methods* include study designs, theory, simulations, and examples related to advanced causal inference methodology with potential EDM applications. *Problems and Perspectives* evaluate the current state of casual inference in EDM, pose areas in which research should be more robust, and describe future directions for the field.

Based on these three guiding themes, we will solicit work on topics including, but not limited to:

- Experimental Designs (e.g., RCTs, A/B testing)
- Quasi-experimental Methods (e.g., Propensity Score Matching, Regression Dissociation Design)
- Effect Estimation (e.g., LOOP, reLOOP, Doubly Robust PSM)
- Directed acyclic graph
- Causal mechanism/mediation analysis
- Effect heterogeneity/moderation annalists
- Machine learning and causal inference
- Principal stratification

3. SCHEDULE AND FORMAT

The proposed schedule is provided in Table 1 and key sections are described below.

3.1 CausalEDM: Retrospective and Future Directions

We will open with a presentation of an empirical retrospective causal work conducted by the EDM community that will go over themes and focuses of past work. This presentation will also include a critical perspective on the needs and gaps for causal work in EDM and propose new ways to evaluate the complex causal connections between innovations in digital technologies and students' learning processes and outcomes. We aim for this presentation to provide an introduction to the topic for graduate students and researchers with other backgrounds as well as to set the stage for the workshop themes throughout the day.

3.2 Panel Mediated Discussion

The *Retrospective and Future Directions* presentation will provide a starting place for the panel-mediated discussion, which will focus on future directions for causal EDM work. The panel will consist of causal inference experts, creators of A/B testing platforms, and prominent EDM scholars who do not focus on causal work but may provide important viewpoints on the needs and directions of this work in EDM. The organizers will facilitate the discussion, and all attendees will be encouraged to participate.

3.3 Paper Presentations/ Lighting Talks

For the second half of the workshop, researchers will present papers and lightning talks of causal work in EDM. This work will focus on one of the three themes (Empirical Studies, Methods, Problems and Perspectives). We will specifically solicit work that connects recent innovations in instructional designs to students' learning processes and outcomes. Each presentation will end with a moderated question-and-answer session.

3.3.1 Short Papers (6 pages)

Short papers will present partially or fully completed work on one of the three themes. This work must have a casual element and be related to EDM. Previously published work is acceptable. This work will not be published in the EDM proceedings, but upon the author's request, we will make the papers available on the workshop website. Authors will have approximately 10 minutes to present.

3.3.2 Lightning Talks (500 word abstract)

Lightning Talks will present ideas or partially completed work that focuses on one of the three themes. Study designs are welcome. These talks can include unresolved causal problems the field should address. These talks are meant to open up discussions for feedback and thought. We encourage graduate students and early-career researchers to consider sharing in-progress research in this format to solicit feedback and foster future collaborations.

3.3.3 Reviewing Process

The workshop organizers will review papers and abstracts alongside ad hoc external reviewers whose expertise is appropriate for the submissions.

3.4 Closing Discussion

Finally, we will conclude with a list of causal problems and opportunities for EDM. This discussion will be open to attendees and panel participants.

4. PREVIOUS EDITIONS AND EXPECTED PARTICIPANTS

With previous workshops held annually between 2020 and 2024, this convening would represent the fifth edition of this workshop at EDM. Over 30 individuals attended each iteration, and the EDM community's interest in causal inference

topics has been steadfast. We aim to resume this convening at EDM 2025 and build on participation from previous iterations by providing entry points for researchers who may not typically use causal methods. For example, our opening presentation on causal inference methods in EDM provides an introduction for those who have not previously participated in CausalEDm to gain an overview of the work. The lightning talks also provide opportunities for researchers who are considering or beginning to work on causal problems to present work and ideas. With the co-location and similar timing of the Artificial Intelligence in Education (AIED) and Learning @ Scale (L@S) conferences, as well as recent initiatives promoting research in digital learning platforms including SEERNet[3] and the AIMS Colaboratory[4], we anticipate similar levels of high interest and attendance for this workshop.

5. ORGANIZERS

Kirk Vanacore (University of Pennsylvania). Kirk is a Postdoctoral Researcher at the Penn Center for Learning Analytics. He applies statistical and machine learning models to data from computer-based learning platforms to understand learning mechanisms. His work includes using experimental designs and observational studies to understand the nuances of how educational programs and pedagogies impact learning. He focuses on the intersection of failure, struggle, and learning.

Anthony F. Botelho (University of Florida). Anthony is an Assistant Professor of Educational Technology and Computer Science Education in the College of Education at the University of Florida. His research seeks to impact learning through the blending of learning theory and quantitative methods. Anthony's primary lines of research include the study of student cognition, behavior, and affect, identifying effective learning interventions through causal inference, and developing human-in-the-loop systems and tools to support teachers.

Avery H. Closser (University of Florida). Avery is an Assistant Professor of Emerging Technologies and Learning in the College of Education at the University of Florida. Her research leverages cognitive theory to advance learning technologies and the design of open materials for instructional practice. She specializes in experimental design in the context of learning technologies and explores best practices for methodologies related to this area of research.

Adam C. Sales (Worcester Polytechnic Institute). Adam is an Assistant Professor of Mathematical Sciences and an affiliate of the Learning Sciences and Technologies and Data Science programs at WPI. His research in applied statistics focuses on methods for causal inference using large, administrative datasets, primarily with applications in learning sciences and social sciences. He has developed and worked on methods combining machine learning with design-based analysis of randomized trials and matched observational studies, principal stratification and mediation analysis using log data from intelligent tutoring systems, and regression discontinuity designs.

[3]seernet.org
[4]aimscollaboratory.org

Neil T. Heffernan (Worcester Polytechnic Institute). Neil is the William Smith Dean's Professor of Computer Science at WPI, the creator of ASSISTments, and an active researcher in the fields of 1) artificial intelligence and education, 2) educational data mining and 3) learning analytics. In order to support research in these fields, Dr. Heffernan created the E-TRIALS Testbed, a tool that allows ASSISTments to be used as a platform to do science and support evidence-based practice. He has dozens of papers in educational data mining, and 20+ papers in comparing different ways to optimize student learning.

6. REFERENCES

[1] N. Chakraborty, S. Roy, W. L. Leite, M. K. S. Faradonbeh, and G. Michailidis. The effects of a personalized recommendation system on students' high-stakes achievement scores: A field experiment. *International Educational Data Mining Society*, 2021.

[2] A. H. Closser, A. Botelho, J. Chan, et al. Exploring the impact of symbol spacing and problem sequencing on arithmetic performance: An educational data mining approach. *Journal of Educational Data Mining*, 16(1):84–111, 2024.

[3] A. H. Closser, A. Sales, and A. F. Botelho. Should we account for classrooms? analyzing online experimental data with student-level randomization. *Educational technology research and development*, 72:1–30, 2024.

[4] W. Dai, Y.-S. Tsai, J. Lin, A. Aldino, H. Jin, T. Li, D. Gašević, and G. Chen. Assessing the proficiency of large language models in automatic feedback generation: An evaluation study. *Computers and Education: Artificial Intelligence*, 7, Dec. 2024.

[5] W. Dai, Y.-S. Tsai, J. Lin, A. Aldino, H. Jin, T. Li, D. Gašević, and G. Chen. Assessing the proficiency of large language models in automatic feedback generation: An evaluation study. *Computers and Education: Artificial Intelligence*, 7:100299, Dec. 2024.

[6] Y. Fan, L. Tang, H. Le, K. Shen, S. Tan, Y. Zhao, Y. Shen, X. Li, and D. Gašević. Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. page bjet.13544, Dec. 2024.

[7] W. L. Leite, Z. Jing, H. Kuang, D. Kim, and A. C. Huggins-Manley. Multilevel mixture modeling with propensity score weights for quasi-experimental evaluation of virtual learning environments. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(6):964–982, 2021.

[8] W. L. Leite, H. Kuang, Z. Shen, N. Chakraborty, G. Michailidis, S. D'Mello, and W. Xing. Heterogeneity of treatment effects of a video recommendation system for algebra. In *Proceedings of the Ninth ACM Conference on Learning@ Scale*, pages 12–23, 2022.

[9] B. A. Motz, P. F. Carvalho, J. R. de Leeuw, and R. L. Goldstone. Embedding experiments: Staking causal inference in authentic educational contexts. *Journal of Learning Analytics*, 5(2):47–59, 2018.

[10] M. Pankiewicz and R. S. Baker. Navigating compiler errors with ai assistance - a study of gpt hints in an introductory programming course. In *Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1*, page 94–100, Milan Italy, July 2024. ACM.

[11] Z. A. Pardos and S. Bhandari. Chatgpt-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills. *PLOS ONE*, 19(5):e0304013, May 2024.

[12] D. M. Pham, K. P. Vanacore, A. C. Sales, and J. A. Gagnon-Bartsch. Lool: Towards personalization with flexible robust estimation of heterogeneous treatment effects. In *Proceedings of the 17th International Conference on Educational Data Mining*, pages 376–384. International Educational Data Mining Society, July 2024.

[13] E. Prihar, M. Syed, K. Ostrow, S. Shaw, A. Sales, and N. Heffernan. Exploring common trends in online educational experiments. In *Proceedings of the 15th International Conference on Educational Data Mining*, pages 27–38, 2022.

[14] A. C. Sales and J. F. Pane. The role of mastery learning in an intelligent tutoring system: Principal stratification on a latent variable. *The Annals of Applied Statistics*, 13(1):420–443, Mar. 2019.

[15] A. C. Sales, E. B. Prihar, J. A. Gagnon-Bartsch, and N. T. Heffernan. Using auxiliary data to boost precision in the analysis of a/b tests on an online educational platform: New data and new results. *Journal of Educational Data Mining*, 15(2):53–85, 2023.

[16] A. C. Sales, K. P. Vanacore, H.-A. Kang, and T. A. Whittaker. Problem-solving behavior and edtech effectiveness: A model for exploratory causal analysis. In *Proceedings of the 17th International Conference on Educational Data Mining*, pages 385–395. International Educational Data Mining Society, July 2024.

[17] K. Vanacore, M. Pankiewicz, and R. Baker. Unpacking the impact of generative ai feedback: Divergent effects on student performance and self-regulated learning. Feb 2025.

[18] K. P. Vanacore, A. Gurung, A. A. McReynolds, A. Liu, S. T. Shaw, and N. T. Heffernan. Impact of non-cognitive interventions on student learning behaviors and outcomes: An analysis of seven large-scale experimental inventions. In *LAK23: 13th International Learning Analytics and Knowledge Conference*, page 10, 2023.

[19] K. P. Vanacore, A. Gurung, A. Sales, and N. Heffernan. Effect of gamification on gamers: Evaluating interventions for students who game the system: Evaluating interventions for students who gaming the system. *Journal of Educational Data Mining*, 16(11):112–140, June 2024.