Jiayi (Joyce) Zhang

+1 (412)-802-5175 | jzhang31@wpi.edu

LinkedIn | Github | Google Scholar

Philadelphia, PA , United States

Bio

I am a post-doctoral researcher in the Computer Science Department at Worcester Polytechnic Institute, working with Dr. Neil Heffernan. Prior to this, I obtained my doctoral degree in Learning Sciences and Technologies from the University of Pennsylvania, where I was mentored by Dr. Ryan Baker.

My research investigates how artificial intelligence can be leveraged efficiently and responsibly to facilitate learning through data-driven approaches. I specialize in developing scalable and robust methods that model and understand students' complex learning behaviors and processes, including affect, engagement, and self-regulated learning strategies. My work has been published in leading international conferences and peer-reviewed journals and has received multiple research awards.

Research interests: AI in Education, Learning Analytics, Educational Data Mining, Student Modeling, Self-Regulated Learning, AI Ethics

Education

- **University of Pennsylvania,** Philadelphia, PA — *Sept 2020 - Dec 2025*
  Ph.D., Learning Sciences and Technologies
  - Dissertation topic: Measuring and Understanding Students' Self-Regulated Learning in Textual Data in Computer-Based Learning Environments
- **Teachers College, Columbia University,** New York, NY — *Sept 2015 – May 2017*
  M.A., Cognitive Science in Education
- **University of Pittsburgh,** Pittsburgh, PA — *Sept 2011 – May 2015*
  B.S., Psychology & Economics

Professional Experience

- **Post-Doctoral Fellow/ Research Scientist.** ASSISTments, Worcester Polytechnic Institute [🌐] — *Jan 2026 - present*
- **Graduate Research Fellow.** Penn Center for Learning Analytics, University of Pennsylvania [🌐] — *Sept 2020 - Dec 2025*
- **Educational Research Analyst.** Imagine Learning [🌐] — *Nov 2017 - Aug 2020*
- **Research Intern.** Amplify Education [🌐] — *July - Oct 2017*
- **Research Intern.** Macmillan Learning [🌐] — *Sept - Dec 2016*
- **Intern.** Wildlife Conservation Society [🌐] — *May - Aug 2016*
- **Research Assistant.** Peer Assessment Study, University of Pittsburgh [🌐] — *May 2013 - May 2015*
- **Research Assistant.** Parents Anxiety, University of Pittsburgh — *May 2013 - Sept 2013*

Teaching Experience

- **Co-Instructor.** EDUC 6195. Capstone Seminar: Learning Analytics. University of Pennsylvania — *June - Aug 2026*
- **Co-Instructor.** EDUC 6191. Core Methods in Educational Data Mining. University of Pennsylvania — *Sept - Dec 2024*
- **Guest Lecturer.** EDIS 8100. Teaching and Learning Analytics. University of Virginia
  - Text-Based Analytics and Natural Language Processing — *Sept, 2026*
- **Guest Lecturer.** CUIN 7397. Research in the Age of AI. University of Houston
  - Applications of GenAI in Interacting with Scholarly Literature — *Sept, 2026*
- **Guest Lecturer.** CS565. User Modeling. Worcester Polytechnic Institute
  - LLM for Student Modeling. — *Nov, 2025*
- **Guest Lecturer.** EDUC6191. Core Methods in Educational Data Mining. University of Pennsylvania
  - Bayesian Knowledge Tracing. — *Oct 2023, 2025*
- **Teaching Assistant.** Introduction to Psychology. University of Pittsburgh — *Sept - Dec 2014*

Grants, Fellowships, and Scholarships

- **National Science Foundation.** Curating a Large-Scale Publicly Available Multimodal Dataset of Student Actions in Illustrative Math to Study Mathematical Misconceptions. 2026-2029. Heffernan, N. (PI), Zhang, J. (Co-PI). $1.5M. Award #: 2621434 [🌐]
- **Dean's Fellowship.** University of Pennsylvania. 2020-2025.
- **Ethical AI in Education Fellowship.** Learnest. 2024. [🌐]

Research Awards

- **Best Student Paper.** International Conference on Artificial Intelligence in Education (AIED). 2026
- **Best Presentation.** International Conference on Computer Science and Technologies in Education (CSTE). 2025
- **Best Student Paper.** International Conference on Quantitative Ethnography (ICQE). 2024
- **Best Paper Nomination.** Learning Analytics and Knowledge Conference (LAK). 2024
- **Best Paper Nomination.** International Conference on Educational Data Mining (EDM). 2022
- **Best Poster Presentation Runner-up.** International Conference on Educational Data Mining (EDM). 2021
- **Best Paper Nomination.** International Conference on Artificial Intelligence in Education (AIED). 2021

Publications

C=Conference, J=Journal

Citation Metrics

Number of publications: 50 *(Google Scholar, September 27th, 2026)*
Total citations: 748 *(Google Scholar, September 27th, 2026)*
h-index: 12 *(Google Scholar, September 27th, 2026)*
i10-index: 18 *(Google Scholar, September 27th, 2026)*
Number of distinct co-authors on listed papers: 114

Journal Articles

[J.5] Zambrano, A., Wei, Z., **Zhang, J.**, Baker, R., Ocumpaugh, J., Barany, A., Liu, X., Zhou, Y., Paquette, L., Ginger, J., & Borchers, C. (2025) Data Plus Theory Equals Codebook: Leveraging LLMs for Human-AI Codebook Development. *Journal of Educational Data Mining*.

[J.4] Borchers, C., **Zhang, J.**, Fleischer, H., Schanze, S., Aleven, V., & Baker, R. S. (2025). Large Language Models Generalize SRL Prediction to New Languages Within But Not Between Domains. *Journal of Educational Data Mining*.

[J.3] Liu, X., Zambrano, A. F., Baker, R. S., Barany, A., Ocumpaugh, J., **Zhang, J.**, Pankiewicz, M., Nasiar, N., & Wei, Z. (2025). Qualitative Coding with GPT-4: Where It Works Better. *Journal of Learning Analytics*, 12(1), 169–185.

[J.2] Baker, R. S., Richey, J. E., **Zhang, J.**, Karumbaiah, S., Andres-Bray, J. M., Nguyen, H. A., Andres, J. M. A. L., & McLaren, B. M. (2024). Gaming the system mediates the relationship between gender and learning outcomes in a digital learning game. *Instructional Science*, 1–36.

[J.1] **Zhang, J.**, Andres, J. M. A. L., Hutt, S., Baker, R. S., Ocumpaugh, J., Nasiar, N., ... & Young, T. (2022). Using Machine Learning to Detect SMART Model Cognitive Operations in Mathematical Problem-Solving Process. *Journal of Educational Data Mining*, 14(3), 76–108.

Conference Papers

[C.45] **Zhang, J.**, Borchers, C., Wang, C., Kumar, V., Teffera, L., McLaren, B., & Baker, R.S. (2026). Understanding Gaming the System through Analyzing Self-Regulated Learning in Think-Aloud Protocols. In *Proceedings of the Learning Analytics and Knowledge Conference (LAK).*

[C.44] **Zhang, J.**, Borchers, C., Cohn, C., Srivastava, N., Snyder, C., Guo, S., T S, A., Mohammed, N., Noh, H., & Biswas, G. (2026). Using Large Language Models to Detect Socially Shared Regulation of Learning in Collaborative Learning. In *Proceedings of the Learning Analytics and Knowledge Conference (LAK).*

[C.43] **Zhang, J.**, Baker, R. S., Mallick, D. B., Heffernan, C., & Heffernan, N. (2026). A Case Study Reexamining the Cold-Start Problem in Knowledge Tracing Models and Implications for SafeInsights, an Education Research Infrastructure. The Seventh Annual Workshop on A/B Testing and Platform-Enabled Learning Engineering (PELE) at Learning@Scale. arXiv preprint arXiv:2606.11004.

[C.42] Zambrano, A., **Zhang, J.**, Pankiewicz, M., & Baker, R. S. (2026). Practice as the Key to Success: Understanding the Role of Prior Knowledge, Affective States and Learning Resources in Computer Science Education. In *Proceedings of the Learning Analytics and Knowledge Conference (LAK).*

[C.41] Wu, M., **Zhang, J.**, Zhang, R. Z. (2026). Decoding Human and AI Persuasion in National College Debate: Analyzing Prepared Arguments Through Aristotle's Rhetorical Principles. *American Educational Research Association (AERA).*

[C.40] Xiao, F., **Zhang, J.**, Zambrano, A., & Jiang, S. (2026). What Children's AI Literacy Books Teach About Artificial Intelligence: A Content Analysis Using the AI4K12 Framework. *Proceedings of the International Conference on Artificial Intelligence in Education (AIED).* **[Won Best Student Paper Award]**

[C.39] Borchers, C., **Zhang, J.**, & Gurung, A. (2026). Representation Learning to Study Temporal Dynamics in Tutorial Scaffolding. *Proceedings of the International Conference on Artificial Intelligence in Education (AIED).*

[C.38] Huang, X., Borchers, C., **Zhang, J.**, & Lajoie, S. P. (2026). Physiological and Semantic Patterns in Medical Teams Using an Intelligent Tutoring System. *Proceedings of the International Conference on Artificial Intelligence in Education (AIED).*

[C.37] Gao. J., **Zhang, J.,** & Chen. D., (2026). To Use or Not to Use: Investigating Student Perceptions of Faculty Generative AI Usage in Higher Education. *Proceedings of the International Conference on Artificial Intelligence in Education (AIED).*

[C.36] Worden, E., Lee, M., Siedahmed, A., Sales, A., **Zhang, J.**, Shraga, R., & Heffernan, N. (2026). Short, long, or affective: Evaluating LLM-generated feedback styles for student learning. *Proceedings of the International Conference on Artificial Intelligence in Education (AIED).*

[C.35] Deho, O., Baker, R., Smith, D., **Zhang, J.**, & Woodhead, S. (2026). To what extent can early keystroke patterns predict knowledge gain in mathematics. *Proceedings of the International Conference on Educational Data Mining (EDM).*

[C.34] Worden, E., Dang, L., Lim, W. C., Miller, S., **Zhang, J.**, Haim, A., Sales, A., Gurung, A., & Heffernan, N. (2026). A large-scale randomized controlled trial showing LLM-generated feedback helps low-knowledge middle school math students with short-term learning. ACM Conference on Learning@Scale (L@S).

[C.33] Lee, S., Zhang, S., **Zhang, J.**, Zambrano, A., Li, H., Adjei, S., Baker, R.S., Botelho, A. (2026). A Tale of Two Models: Leveraging Self-Explanation to Contextualize Divergent Models of Student Carelessness. OSF.

[C.32] Wei, Z., **Zhang, J.**, Baker, R.S., Borchers, C., Barany, A., Liu, X. (2026) Scaffolding Human-LLM Qualitative Coding Through Inference Evaluation and Construct Boundaries. *Proceedings of the International Conference on Quantitative Ethnography (ICQE).*

[C.31] **Zhang, J.**, Baker, R. S., Srivastava, N., Ocumpaugh, J., Mills, C., & McLaren, B. M. (2025). Carelessness detection using performance factor analysis: A new operationalization with unexpectedly different relationship to learning. In *Proceedings of the International Conference on Computer Science and Technologies in Education*. **[Best Presentation]**

[C.30] **Zhang, J.**, Vanacore, K., Baker, R., Ch, N., Mills, C., & Henkel, O. (2025). How much mastery is enough mastery? The relationship between mastery in a lesson and the performance on the subsequent lesson. In *Proceedings of the International Conference on Educational Data Mining (EDM).*

[C.29] **Zhang, J.**, Baker, R.S., & McLaren, B.M. (2025). LLM-Based Approaches for Detecting Gaming the System in Self-Explanation. In *Proceedings of the Artificial Intelligence in Measurement and Education Conference (AIME-Con).*

[C.28] **Zhang, J.** (2025). Examining the Robustness of Large Language Models across Language Complexity. *American Educational Research Association (AERA).*.

[C.27] Liu, X., Wei, Z., Baker, R. S., Metcalf, S. J., **Zhang, J.**, Barany, A., Slater, S., Swanson, L., & Gagon, D. (2025). Integrating Large Language Models and Machine Learning to Detect Struggle in Educational Games. In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED).*

[C.26] Liu, Y., Davidesco, I., McLaren, B., Richey, E., Xue, X., Teffera, L., Stec, H., **Zhang, J.**, Liu, X., & Golumbic, E. (2025). Utilizing Log-based and Neurophysiological Measures to Understand Engagement and Learning with Intelligent Tutoring Systems. In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED).*

[C.25] Moreau-Pernet, B., Vanacore, K., **Zhang, J.**, Baker, R., & Henkel, O. (2025). Keeping Students in their Zone of Proximal Development: A Machine Learning Approach for Intelligent Tutoring Systems. OSF

[C.24] **Zhang, J.**, Borchers, C., Aleven, V., & Baker, R. S. (2024). Using large language models to detect self-regulated learning in think-aloud protocols. In *Proceedings of the International Conference on Educational Data Mining (EDM).*

[C.23] **Zhang, J.**, Borchers, C., & Barany, A. (2024). Studying the Interplay of Self-Regulated Learning Cycles and Scaffolding Through Ordered Network Analysis Across Three Tutoring Systems. In *Proceedings of the International Conference on Quantitative Ethnography (ICQE).***[Won Best Student Paper Award]**

[C.22] Liu, X., **Zhang, J.**, Barany, A., Pankiewicz, M., & Baker, R.S. (2024). Assessing the Potential and Limits of Large Language Models in Qualitative Coding. In *Proceedings of the International Conference on Quantitative Ethnography (ICQE).*

[C.21] Borchers, C., **Zhang, J.**, Baker, R. S., & Aleven, V. (2024). Using Think-Aloud Data to Understand Relations between Self-Regulation Cycle Characteristics and Student Performance in Intelligent Tutoring Systems. In *Proceedings of the Learning Analytics and Knowledge Conference (LAK).*

[C.20] Zambrano, A. F., **Zhang, J.**, & Baker, R. S. (2024). Investigating Algorithmic Bias on Bayesian Knowledge Tracing and Carelessness Detectors. In *Proceedings of the Learning Analytics and Knowledge Conference (LAK).***[Nominated for Best Paper Award]**

[C.19] Cloude, E. B., **Zhang, J.**, Baker, R. S., & Fouh, E. (2024). Procrastination vs. Active Delay: How Students Prepare to Code in Introductory Programming. In *Proceedings of the ACM Technical Symposium on Computer Science Education*

[C.18] Švábenský, V., Pankiewicz, M., **Zhang, J.**, Cloude, E. B., Baker, R. S., & Fouh, E. (2024). Comparison of Three Programming Error Measures for Explaining Variability in CS1 Grades. In *Proceedings of the Innovation and Technology in Computer Science Education*.

[C.17] Borchers, C., Liu, X., Lee, H. H., & **Zhang, J.** (2024). Ethical AIED and AIED Ethics: Toward Synergy Between AIED Research and Ethical Frameworks. In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED).*

[C.16] Zambrano, A., Ocumpaugh, J., Nasiar, N., Goslen, A., **Zhang, J.**, Rowe, J., ... & Hutt, S. (2024). Says Who? How different ground truth measures of emotion impact student affective modeling. In *Proceedings of the International Conference on Educational Data Mining (EDM).*

[C.15] Barany, A., Nasiar, N., Porter, C., Zambrano, A. F., Andres, A. L., Bright, D., ... & Baker, R. S. (2024). ChatGPT for education research: Exploring the potential of large language models for qualitative codebook development. In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED).*

[C.14] **Zhang, J.**, Baker, R. S., Andres, J. A., Hutt, S., & Sethuraman, S. (2023). Automated Multi-Dimensional Analysis of Peer Feedback in Middle School Mathematics. In *Proceedings of the International Society of the Learning Sciences (ISLS).*

[C.13] **Zhang, J.**, Pang, S., Andres, J. A., Baker, R. S., Cloude, E., Anh, H., & Nguyen, B. M. (2023). Leveraging Natural Language Processing to Detect Gaming the System in Open-ended Questions in a Math Digital Learning Game. Presented at the Annual Meeting of the Society for Text and Discourse.

[C.12] **Zhang, J.**, Baker, R. S., & Farmer, T. (2023). No Benefit for High-Dosage Time Management Interventions in Online Courses. In *Proceedings of the ACM Conference on Learning@ Scale*.

[C.11] Nasiar, N., Baker, R. S., Zou, Y., **Zhang, J.**, & Hutt, S. (2023). Modeling problem-solving strategy invention (PSSI) behavior in an online math environment. In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED).*

[C.10] Goslen, A., Henderson, N., Rowe, J., **Zhang, J.**, Hutt, S., Ocumpaugh, J., ... & Lester, J. (2023). Enhancing Engagement Modeling in Game-Based Learning Environments with Student-Agent Discourse Analysis. In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED).*

[C.9] Andres, J. M., Baker, R. S., Hutt, S. J., Mills, C., **Zhang, J.**, Rhodes, S., & DePiro, A. (2023). Anxiety, achievement, and self-regulated learning in CueThink. In *Proceedings of the International Society of the Learning Sciences (ISLS).*

[C.8] **Zhang, J.**, Andres, J. M. A. L., Hutt, S., Baker, R. S., Ocumpaugh, J., Mills, C., Brooks, J., Sethuraman, S., & Young, T. (2022). Detecting SMART Model Cognitive Operations in Mathematical Problem-Solving Process. In *Proceedings of the International Conference on Educational Data Mining (EDM).***[Nominated for Best Paper Award]**

[C.7] **Zhang, J.**, Hutt, S., Ocumpaugh, J., Henderson, N., Goslen, A., Rowe, J. P., ... & Lester, J. (2022). Investigating Student Interest and Engagement in Game-Based Learning Environments. In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED).*

[C.6] **Zhang, J.**, Cunningham, T., Iyer, R., Baker, R., & Fouh, E. (2022). Exploring the Impact of Voluntary Practice and Procrastination in an Introductory Programming Course. In *Proceedings of the ACM Technical Symposium on Computer Science Education*.

[C.5] Karumbaiah, S., **Zhang, J.**, Baker, R. S., Scruggs, R., Cade, W., Clements, M., & Lin, S. (2022). Using Neural Network-Based Knowledge Tracing for a Learning System with Unreliable Skill Tags. In *Proceedings of the International Conference on Educational Data Mining (EDM).*

[C.4] Wu, M., **Zhang, J.**, & Barany, A. (2022). Understanding Detectors for SMART Model Cognitive Operation in Mathematical Problem-Solving Process: An Epistemic Network Analysis. In *Proceedings of the International Conference on Quantitative Ethnography (ICQE).*

[C.3] He, M., Baker, R., Hutt, S., & **Zhang, J.** (2022). A Less Overconservative Method for Reliability Estimation for Cohen's Kappa.

[C.2] **Zhang, J.**, Das, R., Baker, R. S., & Scruggs, R. (2021). Knowledge Tracing Models' Predictive Performance when a Student Starts a Skill. In *Proceedings of the International Conference on Educational Data Mining (EDM).***[Runner-up for Best Poster Presentation]**

[C.1] Richey, J. E., **Zhang, J.**, Das, R., Andres-Bray, J. M., Scruggs, R., Mogessie, M., ... & McLaren, B. M. (2021). Gaming and confrustion explain learning advantages for a math digital learning game. In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED).***[Nominated for Best Paper Award]**

Invited Talks

- **Learning How to Learn in the Age of AI** Workshop at EPFL - Swiss Federal Technology Institute of Lausanne. October 2025
- **Examining the robustness of Large Language Models based on Language Complexity.** AI Ethics Colloquium at Stanford University. May 2024
- **Leveraging NLP to Detect Gaming the System in Open-ended Questions in a Math Digital Learning Game.** Natural Language Processing workshot at Learning Analytics and Knowledge Conference (LAK). March 2023

Journal, Conference, Grant Reviews

- National Science Foundation (NSF) Panelist. 2026
- International Learning Analytics and Knowledge Conference (LAK). 2025
- International Conference on Artificial Intelligence in Education (AIED). 2025, 2026
- International Conference on Educational Data Mining (EDM). 2024, 2025, 2026
- ACM Learning @ Scale Conference (L@S). 2026
- International Conference on Quantitative Ethnography (ICQE). 2024, 2025
- Artificial Intelligence in Measurement and Education Conference (NCME AIME-Con). 2025
- Computers & Education
- Learning and Individual Differences
- Journal of Educational Data Mining
- Computer-Based Learning in Context Journal
- IEEE Transactions on Learning Technologies

Leadership and appointments

- **Publication Co-Chair.** International Conference on Educational Data Mining (EDM). 2027

Mentorship

- **Mentor.** Teachers College Alumni Mentorship Program, Teachers College, Columbia University, 2026–2027
- **Mentor.** Learning Analytics and Artificial Intelligence Capstone Project, University of Pennsylvania, 2026-2027
- **Ph.D. students mentored:** Eamon Worden, Morgan Lee, Abubakir Siedahmed, Ethan Croteau
- **Master's students mentored:** Caroline Buchanan, Wei Liu, Takayoshi Koyama, Sophia Zhao, Shiyi Pang, Tianyue Bao, Xiaodan Yu, Mengqian Wu, Xialin Xie, Zhongtian Huang, Margaret Guan, Yipu (Alison) Zheng
- **Undergraduate and high school students mentored:** Rohini Das, Matt He