XINGYAO (DORIA) XIAO

doria.xiao@acer.org | doriaxiao.github.io | ORCID 0000-0001-8430-0438 | Melbourne, Australia

RESEARCH INTERESTS

Psychometrics and educational measurement; Bayesian latent variable modeling; item response theory (IRT) and Many-Facet Rasch models; growth mixture models and model selection; automated scoring and AI/LLM evaluation; measurement invariance and cross-cultural assessment; integration of machine learning with human judgment systems.

EDUCATION

Ph.D. in Social Research Methodologies, GPA: 4.00/4.00 August 2020 - May 2025
University of California, Berkeley

- Research areas: Bayesian longitudinal and latent variable modeling, growth mixture models, model selection, multidimensional IRT, Many-Facet Rasch Model (MFRM), integration of AI in education

M.S. in Applied Statistics and Psychometrics, GPA: 3.88/4.00 (Major GPA: 4.00) August 2018 - May 2020
Boston College

- Dean's Merit Scholarship (2018 & 2019)

B.A. in Statistics, Minor in Mathematics, GPA: 3.82/4.00 (Major GPA: 4.00) January 2016 - September 2018
University of Minnesota, Morris

- Graduated with Distinction; Project Stipend; Outstanding Performance Award; Dean's List

RESEARCH EXPERIENCE

Research Fellow 2 August 2026 - Present
Australian Council for Educational Research (ACER), Melbourne

- Psychometric methods for large-scale assessment and automated scoring of constructed responses

Postdoctoral Scholar July 2025 - July 2026
LEVANTE Project, Stanford University (Graduate School of Education & Department of Psychology)

- Worked with Ben Domingue, Mike Frank, and Nilam Ram on developmental change and psychometrics research
- Focused on measurement invariance and statistical methods for modeling behavioral and developmental change
- Contributed to research design and engagement within Stanford's psychology and education research communities
- Collaborative project supported by the Jacobs Foundation with international sites

Lead Researcher March 2025 - July 2025
Influence Score Chat Project, The Munathara Initiative

- Designed NLP-based and GPT-assisted scoring/translation evaluation frameworks

Graduate Student Researcher November 2020 - January 2025
Berkeley Evaluation and Assessment Research (BEAR) Center, UC Berkeley

- Developed and validated digital tasks; implemented MFRM and mIRT models; optimized rater and scoring design for educational assessments

Graduate Student Researcher (Bayesian Modeling) November 2020 - January 2025
Education Research using Stan, UC Berkeley

- Created accessible tutorial on multilevel logistic regression using rstanarm (published Stan Case Study)
- Developed code-based and visual diagnostics for identifiability problems in Bayesian growth mixture models; proposed solutions through careful prior selection

Summer Intern June 2023 - August 2023

Edmentum

- Assessed quality of GPT-generated translations using mixed-method algorithmic and human reviews
- Benchmarked AI translations with human translations to evaluate linguistic accuracy and contextual relevance

Statistical Modeling Consultant June 2022 - April 2023

Chinese Academy of Sciences, Institute of Psychology

- Provided guidance on advanced R-based statistical models; contributed to research projects by refining methodologies and enhancing data analysis

Graduate Student Researcher November 2020 - January 2022

California Computer Science Project (CCSP), UC Berkeley

- Helped develop computer science instructional materials in collaboration with educational specialists; contributed to outreach strategies

Graduate Research Assistant August 2019 - July 2020

City Connects: Mary E. Walsh Center for Thriving Children, Boston College

- Assisted in data collection, statistical analysis (regression, multivariate techniques), and report preparation for evaluating educational intervention programs

Data Analyst May 2019 - June 2020

Innovation in Urban Science Education Lab, Boston College

- Conducted and analyzed qualitative data from cognitive interviews for STEM education programs
- Assisted in developing AI Concept Inventories using psychometric analyses

Statistical Research Consultant February 2019 - June 2019

Research Services, Boston College

- Provided statistical analysis and guidance to researchers and faculty

PUBLICATIONS

Peer-Reviewed Journal Articles

Xiao, X., Domingue, B. W., Ram, N., & Frank, M. C. (2026). Recovering accuracy from miscalibrated item banks: Scoring error is fixable, selection error is not. *Educational and Psychological Measurement*. https://doi.org/10.1177/00131644261486854

Xiao, X., & Cheng, Y. (2026). Which algorithm hurts most? Differential associations of task allocation, scheduling, and performance monitoring with worker well-being across 27 EU countries. *International Archives of Occupational and Environmental Health, 99*(8), 54. https://doi.org/10.1007/s00420-026-02239-x

Xiao, X., & Cheng, Y. (2026). What do I know about AI beyond everyday knowledge? Unveiling misconceptions using item response theory analyses and cognitive interviews. *ACM Transactions on Computing Education, 26*(4), 1-28.

Xiao, X., Patz, R. J., & Wilson, M. R. (2026). Revisiting reliability with human and machine learning raters under scoring design and rater configuration in the many-facet Rasch model. *British Journal of Mathematical and Statistical Psychology*. https://doi.org/10.1111/bmsp.70034

Xue, M., **Xiao, X.**, Liu, Y., & Wilson, M. (2026). On the consistency of automatic scoring with large language models. *Educational and Psychological Measurement*. Advance online publication. https://doi.org/10.1177/00131644261418138

Xiao, X., & Cheng, Y. (2026). Gendered pathways to self-efficacy: Moderating and mediating roles of family, school, and sibling contexts in early adolescence. *European Journal of Psychology of Education, 41*(1), 2.

Xiao, X., Rabe-Hesketh, S., & Skrondal, A. (2025). Bayesian identification and estimation of growth mixture models. *Psychometrika*. https://doi.org/10.1017/psy.2025.11

Xue, M., Liu, Y., **Xiao, X.**, & Wilson, M. (2025). Automatic prompt engineering for automatic scoring. *Journal of Educational Measurement*. https://doi.org/10.1111/jedm.70002

Su, B., **Xiao, X.**, Cheng, Y., Liu, C., & Yang, C. (2025). Trajectories of depressive symptom among college students in China during the COVID-19 pandemic: Association with suicidal ideation and insomnia symptoms. *Suicide and Life-Threatening Behavior*. https://doi.org/10.1111/sltb.70051

Cheng, Y., **Xiao, X.**, Jackson, D., Shah, S. A., Abdus-Sabur, F., Hira, A., … Barnett, M. (2025). Competent but anxious smart greenhouse makers: Findings from a physical computing project. *Journal of Science Education and Technology*, 1-19.

Wang, F., Zhu, X., Pi, L., **Xiao, X.**, & Zhang, J. (2024). Patterns of participation and performance at the class level in English online education: A longitudinal cluster analysis. *Education and Information Technologies, 29*, 15595-15619.

Ma, J., Shen, Z., Wang, N., **Xiao, X.**, & Zhang, J. (2023). Developmental differences in children's adaptation to vehicle distance and speed in street-crossing decision-making. *Journal of Safety Research, 88*, 261-274.

Zhang, J., Liu, F., Chen, Z., Yu, Z., **Xiao, X.**, Shi, L., & Guo, Z. (2023). A multi-level analysis on the causes of train-pedestrian collisions in Southwest China 2011-2020. *Accident Analysis & Prevention, 193*, 107332.

Xiao, X., & Cheng, Y. (2021). Movie title keywords: A text mining and exploratory factor analysis of popular movies in the United States and China. *Journal of Risk and Financial Management, 14*(2). https://doi.org/10.3390/jrfm14020068

Kim, J. M., **Xiao, X.**, & Kim, I. (2020). Hollywood movie data analysis by social network analysis and text mining. *International Journal of Electronic Commerce Studies, 11*(1), 75-92.

Kim, J. M., Lee, N., & **Xiao, X.** (2019). Directional dependence between major cities in China based on copula regression on air pollution measurement. *PLoS ONE, 14*(3), e0213148.

Book Chapters & Technical Reports

Xiao, X., Xue, M., & Cheng, Y. (2023). Bayesian partial credit model and its applications in science education. In X. Liu & W. J. Boone (Eds.), *Advances in Applications of Rasch Measurement in Science Education* (Contemporary Trends and Issues in Science Education, Vol. 57). Springer. https://doi.org/10.1007/978-3-031-28776-3_4

Xiao, X., Ji, F., & Rabe-Hesketh, S. (2022). Introduction to multilevel logistic regression using rstanarm. *Stan Case Study*. https://education-stan.github.io/tutorial_glmm.html

Preprints

Xiao, X., Xue, M., & Cheng, Y. (2026). *Enhancing routing fairness in hybrid scoring with group-conditional coverage*. EdArXiv. https://doi.org/10.35542/osf.io/25nwj_v1

Xiao, X., & Cheng, Y. (2026). *How far can a language model carry a systematic review? Preregistered reliability evidence from screening to measurement-quality coding*. MetaArXiv. https://doi.org/10.31222/osf.io/qhy74_v1

Xiao, X., & Cheng, Y. (2026). *Beyond acceptance: A pre-attitudinal "uninformed" stage in teachers' engagement with AI across 53 education systems*. EdArXiv. https://doi.org/10.35542/osf.io/y7awu_v1

Xiao, X., & Rabe-Hesketh, S. (2026). *A parameterization-invariant DIC*. arXiv. https://arxiv.org/abs/2605.27844

Xiao, X., Ulitzsch, E., Zhang, L., Frank, M. C., & Domingue, B. (2026, May 21). *The safety valve: A mixture IRT approach to modeling guessing behavior*. PsyArXiv. https://osf.io/preprints/psyarxiv/7gtwd_v1

Kachergis, G., O'Reilly, F., Braginsky, M., **Xiao, X.**, Lightbody, A. A., Shannon, K. A., … Frank, M. C. (2025, December 17). *Creation and validation of the LEVANTE core tasks: Internationalized measures of learning and development for children ages 5-12 years*. PsyArXiv. https://doi.org/10.31234/osf.io/r4dhw_v1

Under Review / In Preparation

Xiao, X. (invited; in preparation). Neural networks and autoencoders for model selection in growth mixture models. *British Journal of Mathematical and Statistical Psychology* (Special Issue: Using Machine Learning for Statistical Inference).

Xiao, X., Patz, R., Wilson, M., & Cheng, Y. (submitted). STEM confidence, context, and problem-solving.

Li, Z., & **Xiao, X.** (submitted). Which countries are more meritocratic? Individual- and country-level patterns of meritocratic beliefs.

Xue, M., **Xiao, X.**, Liu, Y., & Wilson, M. (submitted). Extracting information from process data using transformers.

Xiao, X., Li, Z., Liu, Y., & Cheng, Y. (submitted). Identifying sensitive periods for the impact of physical abuse on psychopathology symptoms.

PRESENTATIONS

Invited Presentations

Xiao, X., & Cheng, Y. (2027, February). What do I know about AI beyond everyday knowledge? Unveiling misconceptions using item response theory analyses and cognitive interviews. Invited journal paper presentation, ACM SIGCSE Technical Symposium, Sacramento, CA.

Xiao, X. (2025, September 9). When growth mixture models break: Identifiability failures and misleading model evaluation. BEAR Seminar, UC Berkeley, Berkeley, CA.

National Council on Measurement in Education (NCME)

Xiao, X., Patz, R., & Wilson, M. (2025). Integration of machine learning and human rater scores using the many-facet Rasch model. Denver, CO.

Xiao, X., Gochyyev, P., & Wilson, M. (2025). Comparing the difficulty of selected-response and constructed-response items. Denver, CO.

Xiao, X., Razavi, P., & Powers, S. (2024). Bridging bilingual gaps: An in-depth exploration of ChatGPT's transadaptation capabilities. Philadelphia, PA.

Xiao, X., Patz, R., & Wilson, M. (2024). Designing scoring reliability and instructional support into classroom-based math assessments. Philadelphia, PA.

Xiao, X. (2023). Bayesian comparison of growth mixture models. Chicago, IL.

Xiao, X. (2022). Bayesian growth mixture models for classifying and measuring individual trajectories. San Diego, CA.

American Educational Research Association (AERA)

Xiao, X., Rabe-Hesketh, S., & Cheng, Y. (2023). Bayesian comparison of growth mixture models: To better pursue truth. Chicago, IL.

International Meeting of the Psychometric Society (IMPS)

Xiao, X., & Rabe-Hesketh, S. (2025). When model evaluation fails in growth mixture models: Diagnostic and comparative perspective. Minneapolis, MN.

Xiao, X., & Rabe-Hesketh, S. (2023). Bayesian model evaluation and local identifiability for growth mixture models. College Park, MD.

Xiao, X., Ji, F., & Ernst, A. (2023). Sample heterogeneity in dynamic psychological processes. College Park, MD.

Other Conferences

Xiao, X., Rabe-Hesketh, S., & Skrondal, A. (2024). Bayesian approaches to identifiability and estimation of growth mixture models. Royal Statistical Society International Conference, Brighton, UK.

Xiao, X., Rabe-Hesketh, S., & Skrondal, A. (2024). Bayesian identification, estimation, and diagnostic techniques for growth mixture models using Stan. StanCon, Oxford, UK.

Xiao, X. (2023). Bayesian model evaluation using marginal likelihood for growth mixture models. Modern Modeling Methods (M3) Conference, Storrs, CT.

Xiao, X., Li, Y., & Park, Y. (2022). Automating book-to-curriculum mapping using representation learning. Association for Education Finance & Policy, Denver, CO.

TEACHING EXPERIENCE

Graduate Student Instructor, University of California, Berkeley

- Hierarchical and Longitudinal Modeling (Fall 2021, 2022, 2023)
- Data Analysis in Education Research II (Spring 2022, 2023)

Teaching Assistant, Boston College

- Intermediate Statistics (Spring 2020)
- Statistics I (Fall 2019)

Teaching Assistant, University of Minnesota, Morris

- Introduction to Statistics (Fall 2017)

AWARDS & GRANTS

- Dissertation Completion Fellowship ($30,000), Berkeley School of Education, UC Berkeley, 2024-2025
- Continuing Student Fellowship Award ($8,000), Berkeley School of Education, UC Berkeley, 2022-2024
- Psychometric Society Travel Award, IMPS 2023, sponsored by EdAstra Tech
- Barbara White Bequest Competition Award ($3,300), Berkeley School of Education, UC Berkeley, 2021. Project: Using Item Response Theory to critique, develop, and validate concept inventories in STEM education

PROFESSIONAL SERVICE

- Editorial Board Member, Measurement: Interdisciplinary Research and Perspectives (Taylor & Francis), 2025-Present
- Conference Reviewer and Session Chair, National Council on Measurement in Education (NCME), 2024
- Manuscript Reviewer, Measurement: Interdisciplinary Research and Perspectives, May 2023, September 2025
- Manuscript Reviewer, School Psychology Review, September 2025

TECHNICAL SKILLS

Measurement & Modeling: IRT, mIRT, MFRM, CAT, DIF, growth modeling, Bayesian analysis (Stan, Mplus, R)
Data Science & Assessment Tools: R, Python, SAS, SQL, Power BI, Shiny, Qualtrics, NLP/LLM evaluation, digital task development