Wanjing Anya Ma

wanjingm@stanford.edu | 0000-0001-5761-8707 | anyawma.github.io
anyawma | @AnyaWMa | wanjing-anya-ma

Education

2021 – 2026 **Ph.D., Stanford University** Learning Sciences and Technology Design
Advisors: *Jason Yeatman, Nick Haber, Ben Domingue*

Ph.D. Minor, Stanford University Computer Science

2018 – 2019 **M.S., University of Pennsylvania** Learning Sciences and Technologies
Advisor: *Ryan Baker*

2016 – 2018 **B.S., New York University** Computer Science

B.S., New York University Teaching Chemistry 7-12

2014 – 2015 **Boston University** Computer Science

Research Skills

Languages: Python, R, Java, JavaScript, SQL, HTML, CSS, C, and TypeScript
Tools/Frameworks: Google Cloud, Firebase, Vue.js, D3.js, TensorFlow, PyTorch, React, Spark
Quantitative: A/B testing, RCT, survey/test development, longitudinal data analysis, data mining
Qualitative: interviews, focus groups, RITE testing, concept evaluations, diary studies, usability studies

Research Interests

- Psychometric: Computerized Adaptive Testing, Cognitive Diagnostic Assessments, Progress Monitoring
- Efficacy research: Reading Interventions, Personalized Learning, Quantitative Evaluation of AI Tools
- AI Applications: Large-Language Models for Assessments, Natural Language Processing, and Alignment

Research Experiences

2021 – … **Research Assistant at Brain Development and Education Lab, Stanford University**
Project: Rapid Online Assessment of Reading
Mentors: *Jason Yeatman, Nick Haber, Ben Domingue, Mike Frank*

Data Analytics

• Develop an adaptive testing solution using Item Response Theory and Maximum Fisher Information that increased the testing efficiency to 40%.
• Conduct an online randomized controlled trial to evaluate the effect of trial-by-trial feedback on students' test performance and make actionable suggestions to school partners in selecting testing modes.
• Lead psychometric research: apply statistical analysis and data science techniques to solve operational questions: optimize test assembly, inform standard-setting in the score reports, and analyze longitudinal evidence on the validity (predictive and concurrent) and reliability of assessments.
• Work closely with engineers and research partners to improve the quality and user experience of assessments, score reports, and data access.
• Contribute to the authorship of the technical manual and the application for California Reading Difficulties Risk Screener (Approved!).

Research Experiences (continued)

Design and Implementation

• Design, implement, and deploy large-scale online applications to assess foundational reading skills, serving over 50,000 K-12 students across the U.S., Colombia, Brazil, Italy, and Canada.

• Develop an open-source library, jsCAT, enabling real-time, browser-based computerized adaptive testing for broad application in behavioral research.

Human-AI Alignment and Evaluation Projects

• Fine-tuned Llama2 to align AI agents' responses with previous students' responses; designed strategic human-in-the-loop systems to automate item generation and model item difficulty for creating parallel testing forms.

• Created a large dataset to explore how multimodal embeddings (such as CLIP) predict children's vocabulary development.

• Lead a research collaboration with Microsoft Education to evaluate the efficiency of an AI-powered reading tool in promoting students' sentence reading development.

2024 Summer **Ida Lawrence Research Intern at ETS Research Institute**

Project: Automatic Item Generation of Reading Comprehension Items

Mentors: *Michael Flor, Zuowei Wang*

• Conducted literature review on the construct of inference making and computational techniques related to bridging anaphora recognition and resolution.

• Annotated and analyzed the existing item bank to explore the relationship between item types and the item difficulties.

• Designed and implemented an automated workflow utilizing state-of-the-art LLMs for generating test items aligned with targeted inference types, and established rigorous evaluation criteria to ensure quality and consistency.

2018 – 2019 **Research Assistant at Penn Center for Learning Analytics, University of Pennsylvania**

Project: Linguistic Analysis and a Hybrid Human-Automatic Coach for Improving Math Identity

Mentor: *Ryan Baker*

• Built latent semantic spaces to model 5th-graders' math discourse in Reasoning Mind.

• Conducted the stepwise regression to investigate relationships among students' math discourse, learning outcomes, and their math identity.

Research Assistant at RIDDLE Lab, New York University

Project: Mitosis Idea Manager in Web-Based Inquiry Environment (WISE)

Mentor: *Camillia Matuk*

• Led qualitative and quantitative data analysis to investigate students' scientific inquiry.

• Applied topic modeling to build features to evaluate students' science explanation.

• Created visualizations of students' learning trajectory across the learning unit.

Professional Experiences

2024 **Ida Lawrence Research Intern**, ETS Research Institute, NJ

2019 – 2021 **Chemistry Subject Expert Teacher**, BASIS Independent Brooklyn, NY

Awards and Fellowships

2023 **Distinguished Poster Award**, International Meeting of Psychometrics Society

Stanford Interdisciplinary Graduate Fellowship, Stanford University

2019 **Best Paper Nomination**, International Conference on Computer Supported Collaborative Learning

2018 **Merit-Based Scholarship**, University of Pennsylvania

Letha Hurd Morgan Award, New York University

Honors in Science Education, New York University

Luke Hallenbeck Scholarship, New York University

2017 **John Park Graduate Student Convention Travel Award**, School Science and Mathematics

Undergraduate Student Spotlight, New York University Courant Computer Science

Publications

* indicates equal first-author contributions

Journal Articles

1. **Ma**, **W. A.**, Richie-Halford, A., Burkhardt, A. K., Kanopka, K., Chou, C., Domingue, B. W., & Yeatman, J. D. (2025). Roar-cat: Rapid online assessment of reading ability with computerized adaptive testing. *Behavior Research Methods*, *57*(1), 1–17. https://doi.org/10.3758/s13428-024-02578-y

2. Gijbels, L., Burkhardt, A., **Ma**, **W. A.**, & Yeatman, J. D. (2024). Rapid online assessment of reading and phonological awareness (roar-pa). *Scientific Reports*, *14*(1), 10249. https://www.nature.com/articles/s41598-024-60834-9

3. Yeatman, J. D., Tran, J. E., Burkhardt, A. K., **Ma**, **W. A.**, Mitchell, J., Yablonski, M., Gijbels, L., Townley-Flores, C., & Richie-Halford, A. (2024). Development and validation of a rapid and precise online sentence reading efficiency assessment. *Frontiers in Education*, *9*, 1494431. https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2024.1494431/full

Conference Proceedings

1. Tan, A. W. M., Yu, S., Long, B., **Ma**, **W. A.**, Murray, T., Silverman, R. D., Yeatman, J. D., & Frank, M. C. (2024). Devbench: A multimodal developmental benchmark for language learning. *Advances in Neural Information Processing Systems*. https://doi.org/https://arxiv.org/abs/2406.10215

2. Zelikman, E., *** Ma**, **W. A.**, * Tran, J., Yang, D., Yeatman, J., & Haber, N. (2023). Generating and evaluating tests for k-12 students with language model simulations: A case study on sentence reading efficiency. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2190–2205. https://doi.org/10.18653/v1/2023.emnlp-main.135

3. Matuk, C., **Ma**, **W.**, Sharma, G., & Linn, M. (2019). The lifespan and impact of students' ideas shared during classroom science inquiry. *Proceedings of the 13th Annual International Conference for Computer Supported Collaborative Learning. Lyon: International Society for the Learning Sciences.*, 49–56. https://par.nsf.gov/servlets/purl/10180393

4. **Ma**, **W.** (2017). A computer tool that will allow secondary science teachers to differentiate reading materials for students with varied reading abilities. *Proceedings of the 116th annual convention of the School Science and Mathematics Association*, 14–21. https://www.ssma.org/assets/Proceedings/Proceedings2017FINALWeb.pdf#page=15

Preprints

1. Roncete, K., Klotz, L., **Ma**, **W. A.**, Arteaga, E., Alves, L., Chrispim, R., Diniz, D., Yeatman, J., & Lichand, G. (2025). The opportunities and challenges of digital assessments in low-resource settings: Evidence from measuring reading fluency in brazil. https://doi.org/10.21203/rs.3.rs-5516837/v1

2. Bhat, K. G., Mogan, A. D., Saavedra, A., Fuentes-Jimenez, M., Siebert, J. M., **Ma**, **W. A.**, Townley-Flores, C., Richie-Halford, A., Wilkey, E. D., & Yeatman, J. (2024). Shared and unique influences of phonological processing on reading and math. https://doi.org/10.31219/osf.io/em3bg

3. He-Yueya, J., **Ma**, **W. A.**, Gandhi, K., Domingue, B. W., Brunskill, E., & Goodman, N. D. (2024). Psychometric alignment: Capturing human knowledge distributions via language models. https://arxiv.org/abs/2407.15645

4. Siebert, J. M., Fuentes-Jimenez, M., **Ma**, **W. A.**, Saavedra, A., Townley-Flores, C., & Yeatman, J. (2024). A fair lexical decision task for monolingual and multilingual spanish-speakers. https://osf.io/preprints/psyarxiv/qfdpb

Open Software

1. **Ma**, **W. A.**, Yeatman, J. D., & Richie-Halford, A. (2023). Jscat: Computer adaptive testing in javascript [Open-source software]. https://github.com/yeatmanlab/jsCAT

Selected Presentations

Invited Talks

1. Zelikman, E., *** Ma**, **W. A.**, * Tran, J., Yang, D., Yeatman, J., & Haber, N. Generating and evaluating tests for k-12 students with language model simulations: A case study on sentence reading efficiency. In: HAI: AI+Education Summit: AI in the Service of Teaching Learning. 2024.

Conference Presentations

1. Fuentes-Jimenez, M., **Ma**, **W. A.**, Maximilian, J., Saavedra, A., Townley-Flores, C., Richie-Halford, A., & Yeatman, J. D. Developing a spanish sentence reading efficiency measure fair for multilingual learners: Roar-frase. In: Annual Meeting of the National Council on Measurement in Education [Accepted]. 2025.

2. **Ma**, **W. A.**, & Domingue, B. W. A comparison of the predictive performance of continuous and dichotomous latent trait models. In: Annual Meeting of the National Council on Measurement in Education [Accepted]. 2025.

3. **Ma**, **W. A.**, Fuentes, M., Siebert, J. M., Saavedra, A., Townley-Flores, C., Richie-Halford, A., Domingue, B. W., & Yeatman, J. D. Exploring effects of trial by trial feedback on validity of dyslexia screening. In: Annual Meeting of the National Council on Measurement in Education [Accepted]. 2025.

4. **Ma**, **W. A.**, Zelikman, E., Tran, J. E., Domingue, B. W., Haber, N., & Yeatman, J. D. Developing parallel forms for sentence reading efficiency using llm-based item response simulator. In: Annual Meeting of the National Council on Measurement in Education [Accepted]. 2025.

5. Long, B., **Ma**, **W. A.**, Silverman, R., Yeatman, J., & Frank, M. C. Developmental changes in the precision of visual concept knowledge. In: Vision Science Society. 2024.

6. Tran, J. E., **Ma**, **W. A.**, Burkhardt, A., T., M., Wentzlof, K., Ungashe, A., Fuentes-Jimenez, M., Stone, H., Mitchell, J., Yablonski, M., Gijbels, L., Richie-Halford, A., Townley-Flores, C., & Yeatman, J. D. Improving the efficiency of silent reading measure through timing analyses and automatic ai test generation. In: NCME Special Conference on Classroom Assessment. 2024.

7. **Ma**, **W. A.**, Burkhardt, A. K., & Yeatman, J. D. Exploring parameter invariance for adaptively assessing reading among students with learning differences. In: Annual Meeting of the National Council on Measurement in Education. 2023.

8. **Ma**, **W. A.**, Richie-Halford, A., Burkhardt, A., Kanopka, K., Chou, C., Domingue, B., & Yeatman, J. D. Roar-cat: Rapid online assessment of reading ability with computerized adaptive testing. In: International Meeting of the Psychometric Society. 2023.

9. Tran, J. E., **Ma**, **W. A.**, Gijbels, L., Townley-Flores, C., Siebert, J., Tran, J. E., Murray, T., Fuentes-Jimenez, M., Ramamurthy, M., Richie-Halford, A., & Yeatman, J. D. Rapid online assessment of reading (roar): A platform for developmental cognitive neuroscience research at an unprecedented scale. In: Flux Congress. 2023.

10. **Ma**, **W.**, Kirch, S. A., Sabouri, P., & Zhang, M. Understanding students' dialogic learning experience in an emergent transformative science classroom. In: National Association for Researching Science Teaching Annual International Conference. 2019.

Graduate Teaching Experiences

Stanford University

2024 **EDUC 252**: Introduction to Psychometrics
Teaching Assistant

Professional Activities

Certificates

2018 – 2023 Chemistry Initial Certificate 7–12 with 5–6 Extension, New York State Education Department

Service to Field

Reviewer National Council on Measurement in Education 2023
National Council on Measurement in Education 2024
NeurIPS 2024 Workshop Large Foundation Models for Educational Assessment

Professional Memberships

2023 – · · · · Associations for Computational Linguistics (ACL)
2022 – · · · · National Council on Measurement in Education (NCME)
2023 – · · · · Psychometric Society (IMPS)

Relevant Courses

Stanford University

CS145: Data Management and Data Systems
CS224n: Natural Language Processing with Deep Learning
CS246n: Mining Massive Data Sets
CS247a: Design for Artificial Intelligence
CS229: Machine Learning
CS293: Empowering Educators via Language Technology
CS329x: Human-Centered NLP
CS448b: Data Visualization
PSYCH251: Experimental Design
PSYCH252: Statistical Methods for Behavioral and Social Sciences
PSYCH253: Measurement and the Study of Change in Social Science Research

STATS200: Introduction to Statistical Inference

University of Pennsylvania

CIS 545: Big Data Analytics
CIS 519: Applied Machine Learning