YAWEN MA, PhD Candidate

Center for health informatics, computing and statistics (CHICAS)
Lancaster Medical School, Faculty of Health and Medicine &
Department of Psychology, Faculty of Sciences and Technology
Lancaster University, United Kingdom, LA1 4YW

Email: m.yawen6@gmail.com
GitHub: https://github.com/Yawen-Ma

Overview

I am a final-year PhD researcher in applied statistics with a strong foundation in mathematics and statistics applied to a real-world application (education research). My current focus is on advancing methodological innovations in areas of quantitative psychology and literacy. Through close collaboration with industrial partner Amplify (U.S. based educational technology company), I have developed state of the art expertise in utilisation of large, complex and unstructured datasets complemented by strong skills in engagement and knowledge exchange.

Research expertise and interests: ***longitudinal methods, unstructured complex datasets, item response theory models, unsupervised learning, cluster analysis, latent variable models, cognitive diagnostic models.***

Education

PhD in Statistics and Epidemiology Oct 2021 – April 2024

(with industrial collaborator: Amplify, United States) (expected submission)

Lancaster University, Center for health informatics, computing and statistics (CHICAS)

Lancaster Medical School & Department of Psychology

Funded by Engineering and Physical Sciences Research Council (EPSRC), UK Research and Innovation

Supervisors: Dr. Anastasia Ushakova, Prof. Kate Cain, Dr. Gabriel Wallin

BSc in Mathematics and Applied Mathematics 2017 – 2021

Xiamen University, First Class Honours

Key subjects: Calculus, Algebra, Geometry, Ordinary/Partial Differential Equations, Number Theory, Mathematical Analysis, Real/Complex Analysis, Time Series, Probability, Statistics.

Publications

- Ma, Y., Cain, K. & Ushakova, A. (2024). Application of cluster analysis to identify different reader groups through their engagement with a digital reading supplement. *Computers & Education (Top in Social Science according to Scholar Metrics)*, *214*, 105025. https://doi.org/10.1016/j.compedu.2024.105025.

  Summary: This paper advances literacy field by introducing a novel approach to analyse, structure and classify information from raw log files that sourced from digital reading apps. The key aim of the digital app is to enhance young learners' reading skills and provide individualised reccomendations. Our analyses provided an innovative analytical and Cluster Analysis framework to guide researchers in navigating this novel and complex dataset to achieve this aim.

- Ma, Y. & Teo, L. P. (2022) On Zagier's conjecture about the inverse of a matrix related to double zeta values. *Journal of Integer Sequences*, *25*, Article 22.6.4. https://doi.org/10.48550/arXiv.2106.15260.

  Summary: This paper was developed as a result of my final project for my BSc. We successfully proved a conjecture by the mathematician Zagier regarding the inverse of a matrix, using elementary methods.

In preparation/submitted

- Ma, Y., Cain, K. & Ushakova, A. (2024). Measuring learners' proficiency: Insights into adaptive digital educational environments. Currently in submission. Preregistration: https://osf.io/gu73j.

  Summary: This paper introduces a comprehensive analytical approach based on Item Reponse Theory to assessing performance of reading skills in adaptive digital learning environments, capturing diverse learning

trajectories and generating reliable scores that reflect each learner's unique, mastery-paced path through multiple attempts.

- Ma, Y., Cain, K., Ushakova, A. & Wallin, G. (2024). A statistical framework for dynamic cognitive diagnosis in digital learning environments. Submitted and currently under peer review at the Journal of the Royal Statistical Society: Series A. Preregistration: https://osf.io/nqkub.

  Summary: This study advances cognitive diagnostic assessments by proposing a novel statistical framework that integrates Cognitive Diagnosis Models with longitudinal data, including assessments of multiple skill sets. Our paper enhances the accuracy of attribute mastery evaluations and the assessment of covariate impacts on learning transitions. The applicability of the method is demonstrated through real-world application.

Talks

Oral Conference Presentations

- Ma, Y., Cain, K. & Ushakova, A. (July 2024). *Big data meets the science of reading: Using graphical models to reveal the dynamics between different skills in early literacy development*. Society for The Scientific Study of Reading, Copenhagen, Denmark.
- Ma, Y., Cain, K. & Ushakova, A. (October 2023). *Application of cluster analysis to identify different reader groups through their engagement with a digital reading supplement.* E-LADDA Closing Conference/ELN Pre-Summit Event: Advances in the Study of Language Development and Literacy Learning in the Digital Age, Porto, Portugal.

Poster Conference/Symposium Presentations

- Ma, Y., Cain, K. & Ushakova, A. *Measuring learners' proficiency: Insights into adaptive digital educational environments*.
  Presented at: (April 2024) Postgraduate Symposium, Lancaster, United Kingdom.
- Ma, Y., Cain, K. & Ushakova, A. *Application of cluster analysis to identify different reader groups through their engagement with a digital reading supplement*.
  Presented at: (August 2022) 24th International Conference on Computational Statistics, Bologna, Italy;
  (April 2022) Postgraduate Symposium, Lancaster, United Kingdom.

Knowledge Exchange Presentations

- Amplify (quarterly presentations, 2022-24)
- Invited Seminar (yearly presentations, CHICAS, Lancaster, 2022-24)
- Invited talk at the Workshop: T-READS Tracking Reading and Educational Attainment through Data (led by Prof Cain (Lancaster University) and Prof Roberts (University of Sheffield)

Research, Teaching and Mentoring

Maths and Stats Tutor Sept 2022 - Present

Learning Development, Lancaster University Library, Faculty of Sciences and Technology/Faculty of Health and Medicine

- Leading of the workshop in multiple statistics topics for undergraduate students across the university (field ranging from social sciences to mathematics), enhancing student stats skills.
- Provision of personalized mentoring in one-to-one sessions for undergraduate and postgraduate students, addressing students' specific challenges in various mathematical and statistical areas.
- Feedback from students (average score = 4.5 or 5/5)

Co-Supervision on MSc Health Data Science (with Dr Ushakova and Prof Cain). Summer, 2023

Lancaster Medical School, Lancaster University

- Preparation and organisation of the datasets for the student, clarification of the structure and facilitation of a smooth data analysis of complex data.
- Provision of assistance to the students in identifying relevant research areas and literature for the project.
- Holding weekly meetings with the student.
- Supporting final write up.

Honors and Awards

- Engineering and Physical Sciences Research Council PhD Studentship (~£90k), full fee and stipend (awarded in 2021).
- China Scholarship Council—Outstanding Self-funded Student Award (~$6k) (awarded in 2025).
- Maths and Stats Hub Star (awarded in 2023).

Research Skills

- Programming languages, mathematical packages and other technical skills: R (advanced), python, C, MATLAB, GitHub, LaTeX.
- Competition for the Best R function (Winner). Centre of Health, Informatics, Computing and Statistics, Lancaster University.
- Languages: English, Mandarin (native).

References

Anastasia Ushakova, Assistant Professor in Biostatistics
Lancaster Medical School
Lancaster University
+44(0)7570692041, a.ushakova@lancaster.ac.uk

Kate Cain, Professor
Department of Psychology
Lancaster University
+44(0)1524593990, k.cain@lancaster.ac.uk

Gabriel Wallin, Assistant Professor in Statistics
School of Mathematical Sciences
Lancaster University
+44 (0)7840 348 395, g.wallin@lancaster.ac.uk

Jennifer Zoski, Learning Scientist
Amplify Education
United States
jzoski@amplify.com