Collaborative Research: LogicDS: Fostering Virtual Learning of Data Science Foundations with Mathematical Logic for Rural High School Students

Type of Proposal: (early stage) design and development**. Strand:** learning**.**
Disciplines: mathematics (statistics and data science) and computer science (its integration with data science and real-world problems).
Grade Levels: 9-12

Overview. Data science is an interdisciplinary subject driven by real-world applications and builds on foundations in computing, math, and statistics. It provides excellent opportunities, not only for high school students to prepare their future careers, but also to develop college-ready STEM competencies. However, these opportunities are still inaccessible to the vast majority of students, particularly those from rural communities, because of the complexity of the overlapping disciplines from mathematics, statistics to computer science and the lack of flexible and reliable learning and teaching resources. We propose an innovative approach to integrating the foundations of data science into one course, named as LogicDS using mathematical logic. Mathematical logic provides a language for us to develop precise definitions of statistics concepts to explicate the logic and reasoning underlying these concepts, and thus promote deeper sense making and reasoning. Furthermore, the learning activities will be contextualized in developing computer models to explore real-world problems using datasets of the United Nation Sustainable Development Goals (SDGs). The LogicDS course will be amplified to reach students of rural communities by our partnership with Florida Virtual School (FLVS), a leader in online K-12 education. FLVS's effective online teaching and learning practices, augmented by personalized learning, provide equal opportunities for data science learning for anyone at any time from anywhere.

Intellectual Merit. Lying at the core of our course is a principled, unified, and novel approach naturally and rigorously integrating the foundations of data science by mathematical logic. This approach not only avoids the offering of several separate courses needed by the interdisciplinary nature of data science, but also enhances the students' learning of fundamental concepts in multiple disciplines (e.g., math, statistics and computer science) by explicitly and rigorously connecting them in the context of data-rich real-world applications. In this project, we will develop the LogicDS curriculum, teacher PD, and the supporting online learning system following a design-based research paradigm. We will investigate how the mathematical logic based data science foundation curriculum can be designed to support high school students' understanding of statistics concepts and to foster their computational thinking and we will identify the effective virtual learning strategies to engage rural students, particularly for female students using user research and feasibility study. Furthermore, we will gather preliminary evidence of the efficacy of the LogicDS curriculum using an experimental randomized control trial study with 1000 high school students. We will test the impacts of the LogicDS on students' learning outcomes (understanding of set theory, logic, and statistics concepts, computational thinking, data science practices and their interest in pursuing a data science related career pathway, particularly for rural female students. The findings are anticipated to contribute to the knowledge base on data science learning for rural high school students.

Broader Impact. The work will have important lasting benefits to society. 1. It will better build students' foundation for data science learning, thus acquiring the necessary skills to choose future education and career pathways for workforce development. 2. This project will also broaden the participation of underserved groups in data science learning from rural communities. The lack of diversity in computing and data science fields is well-documented. Reaching and retaining these students is critical for a more equal and diverse society. 3. It will enhance the infrastructure for research and education by developing a set of strategies that can be applied to other subjects and domains to increase interdisciplinary thinking. 4. The work will advance science while promoting teaching, training and learning by providing a way to promote a more effective learning approach to data science learning by bringing in theoretical and practical perspectives on mathematical logic. 5. Broad dissemination will be achieved by publishing research findings at premier conferences and journals in education research, learning sciences, artificial intelligence, computer science and data science. The integrated curriculum and findings about the related learning process will be made freely available to the public through websites and social media channels. We will also leverage the franchise model of FLVS to reach out to more rural communities and beyond.