CAREER: Rethinking How Middle and High School Students Analyze Ecological Data,
Joshua M. Rosenberg, University of Tennessee, Knoxville

Overview: There have been prominent and widespread calls for high school science students to work with data in more complex ways that better align with and support the work of professional scientists and engineers (Lee & Wilkerson, 2018; National Academies of Sciences, Engineering, and Medicine, 2019). This is part of the broader practice shift in science education research and standards and more data-intensive areas of study and work. While teachers express the desire to do more with data in their science classrooms (Banilower et al., 2018; Rosenberg et al., 2022b), science teachers do not presently have a way to connect the work their students do with data to the science ideas they are supporting their students to understand. This is not the case for other science practices: Consider *engaging in argument from evidence*, where a variety of practical and epistemologically-rich strategies help students to structure students' involvement in science practices. The result is that students' work with data can be isolated from the sense-making students are doing about science. In sum, students' investigations with data must be interpreted in a wholesale way, with patterns in data generally, rather than in the context of how confident they are as a result of their investigation in a particular relationship or effect. There is both the need and an opportunity to provide science teachers with practical tools that are grounded in the framework of Bayesian data analytic methods. The specific opportunity for using Bayesian methods involves advances in informal statistical inference, which considers how students not only interpret, but also model data in terms of making probabilistic generalizations from data. This opportunity is supported by current emphases on students investigating phenomena about which they have or can develop interest This project advances high school students' data modeling in the life sciences by taking a Bayesian approach that is supported and studied in terms of an informal statistical inference framework. It is instantiated through a unit, digital tool, and teacher professional development that will be carried out in a way that emphasizes students' data analyses with phenomena of interest, centering on high school standards that relate to ecosystems (*HS-LS2*). The digital tool will build on prior research that supports learners' to update their ideas probabilistically using Bayes' Theorem. It will be built into the *DataClassroom* platform and it will provide students' with functionality to specify their initial ideas and to update them with data entered into DataClassroom for three types of data: continuous, binary, and count. A field experiment will involve comparing students' informal statistical inference using qualitative analyses of written embedded assessments of students' probabilistic generalization from data between 30 teachers using data with their students and 30 using curricular and technological supports for Bayesian methods to interpret and graph data. In addition, changes in teachers' reports about their teaching practices and students' interest will be formatively assessed using an Experience Sampling Method and traditional survey-based approach.

Intellectual merit: A Bayesian approach has several primary benefits. First, cognitive science research points out that students already think about the world in a more or less Bayesian way. Second, Bayesian methods explicitly and directly connect the ideas students are investigating with data by integrating prior ideas in a way that closely aligns with how science teachers use students' ideas in their instruction. Third, there are benefits to treating the results of students' data modeling in science classrooms as intrinsically uncertain---as a way of supporting students to understand the centrality of uncertainty in science. Last, teachers may turn to conventional (frequentist) statistical methods, but these can be challenging for students to use without prior statistics coursework, and the same issue about how statistics, like parameters, confidence intervals, and *p*-values relate to the science idea students are working to understand.

Broader impacts: While many of the science and engineering practices that distinguish recent science education reform efforts have been studied and supported in classrooms by research that considers their knowledge-building roles, analyzing and interpreting data has been investigated less, a condition that is made more concerning given the emphasis on data analytics and data science across disciplines. This work will provide a digital tool and strategies for 60 science teachers and their approximately 1,200 students to support their efforts to model data. The unit and tool will have an open license and they will be shared through workshops serving teachers through DataClassroom and DataNugget's mailing lists and teachers connected with the Great Smoky Mountains Institute and the East TN STEM Hub.

References

Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M. L. (2018). *Report of the 2018 NSSME+*. Horizon Research, Inc.

Lee, V. R., & Wilkerson, M. H. (2018). *Data use by middle and secondary students in the digital age: A status report and future prospects.* https://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_189500.pdf

National Academies of Sciences, Engineering, and Medicine. (2019). Science and engineering for grades 6-12: Investigation and design at the center. National Academies Press.

Rosenberg, J. M., Schultheis, E. H., Kjelvik, M. K., Reedy, A., & Sultana, O. (2022). Big data, big changes? The technologies and sources of data used in science classrooms. *British Journal of Educational Technology,* 1–23. https://doi.org/10.1111/bjet.13245