Automated Techniques for Educational Meta-Analyses to Improve Diversity, Equity and Inclusion of Educational Policy and Classroom Methods

- Rebecca J. Passonneau, College of Engineering
- Brian Belland, College of Education

Abstract: Meta-analysis applied to research on classroom interventions informs educational policy from the district to the national levels, and supports teachers and instructors in informed decision making about instructional methods for diverse student groups. However, the high cost in human effort and elapsed time to perform accurate meta-analysis works against the potential benefits at multiple decision points, including which educational studies to fund based on community needs, as well as the speed at which proven methods could transform educational practice. Current manual methods for meta-analysis produce effect size estimates that are quickly outdated, and which often do not reflect research on underrepresented populations. The proposed effort supports a pilot study on the feasibility of natural language processing (NLP) methods applied to meta-analysis, to test the ability to automatically replicate human coding from a recently completed meta-analysis. The PI, whose research area is NLP, and the co-PI, with numerous published meta-analyses, have an existing collaboration that they seek to leverage in the proposed pilot project. Positive feedback on a previous submission to the NSF Directorate of Education and Human Resources indicated that pilot results would greatly increase the likelihood that a resubmission would lead to an award. The pilot study would address how to create the requisite data, given human coding sheets and corresponding full-texts of journal articles, and how to leverage machine learning methods for high accuracy of automated methods. The pilot study variables pertain to education population and other learner characteristics related to diversity. Outcomes would have significant impact on Penn State's two strategic themes of Transforming Education, and Empowering through Digital Innovation, as well as on theoretical questions in NLP pertaining to semantics and context, and on the explicit and implicit factors that affect human and automated methods for meta-analysis coding variables.