# Foundations for AI-assisted formative assessment feedback for short-answer tasks in large-enrollment classes

Matthew Beckman
Penn State University

13 September, 2022

## Abstract

Research suggests "write-to-learn" tasks improve learning outcomes, yet constructed-response methods of formative assessment become unwieldy with large class sizes. This study evaluates natural language processing algorithms to assist this aim. Six short-answer tasks completed by 1,935 students were scored by several human raters using a detailed rubric and an algorithm. Results indicate substantial inter-rater agreement using quadratic weighted kappa for rater pairs (each QWK > 0.74) and group consensus (Fleiss' Kappa = 0.68). Additionally, intra-rater agreement was estimated for one rater who had scored 178 responses seven years prior (QWK = 0.88). With compelling rater agreement, the study then pilots cluster analysis of response text toward enabling instructors to ascribe meaning to clusters as a means for scalable formative assessment.