Project Summary

Learning to write clear and precise proofs is a challenging but important task for students studying mathematics and computer science. Writing a good proof requires students to internalize a core set of rules and technical skills, while also understanding how to communicate concepts to others.

Students learning to write proofs require timely and reliable feedback so that they can correct their flaws and improve their presentations. The current method of providing this feedback, by having an expert mentor (the instructor or a graduate teaching assistant) grade their assignments manually, is fundamentally unscalable. It becomes problematic for large classes (with enrollments of 200 students or more), and it is completely out of the question for *web-scale* courses, having 10,000 or more participants.

An alternate approach to assessing proofs (both for providing feedback and for grading) is to make use of undergraduate students—either those currently enrolled or those who have already taken the course and are working as undergraduate graders. This *peer assessment* has the advantage that the workforce scales with the number of students enrolled. It also provides students the opportunity to critically review proofs with different levels of quality and correctness, enhancing their own ability to read and write proofs. Peer assessment, of course, poses its own challenge: to ensure both quality and consistency, because the assessment is performed by people with limited mastery of the material and limited practice in critically reviewing proofs. In addition, it is difficult to ensure uniformity across a large number of assessors.

This project will develop and evaluate a system for assessing proofs in undergraduate mathematics and computer science classes via peer evaluation. It approaches the problem as an instance of *human computation*, using computer technology to harness the collective capability of large numbers of people to do useful work. This requires breaking down the task of assessing a proof to make it possible for multiple, nonexpert people to contribute to the assessment. It requires instituting mechanism to ensure the quality, uniformity, and integrity of the assessment process.

This work will build on existing experience with a prototype system that has been used for two years in computer science and discrete mathematics courses at Carnegie Mellon University The current system targets handling large classes using teams of undergraduate graders. The experience to date indicates the feasibility of the basic strategy, but also many aspects that need to be improved to ensure the quality of the assessments and of the learning experience for the assessors. In addition, the research will take initial steps toward an assessment system that can scale to support web-scale courses.

For **intellectual merit**, this research will study a new realm of human computation, harnessing the abilities of students with limited mathematical training to provide reliable and useful assessments of proofs. It will devise structured frameworks in which to analyze proofs, so that assessment can be decomposed into a number of tasks, each of which can be performed by nonexperts. Identifying these structures will also lead to a deeper understanding of how to teach the writing of proofs. In addition, it will explore ways to maximize the learning experience students gain by critically evaluating each others proofs.

As **broader impacts**, this research seeks to enable students in large classes and in web-scale courses to enhance their learning by receiving useful feedback on proofs they write. It will bring to large numbers of students an opportunity that is otherwise available only in environments consisting of small numbers of students getting attention from expert mentors.