We first sought to run a linear regression predicting students' feedback at the beginning of chapters. The column containing this feedback was adjusted such that it was labeled with the previous chapter–the one that students would actually be giving feedback on. We produced a number of these linear models using different configurations of inputs. The model that balanced complexity and accuracy best was using the End-of-Chapter Score (EOC) and the feedback type as inputs, which resulted in an adjusted R^2 of 0.247.

Since the adjusted R^2 value was not changing much, we converted our response variable into a binary to test if we could model more confidently focusing on whether the feedback was generally positive or negative rather than the exact values provided. We created a variable equal to 0 when the feedback score was > 3 and equal to 1 otherwise. We found that using EOC and feedback type alone provided equivalent results to any more complicated model we tried. This model had an accuracy of 80.5%, a precision of 59.4%, and a sensitivity of 67.5%.

However, investigating the odds that the model was producing turned up interesting results. For all four feedback types, higher EOC scores decreased the odds of negative feedback, which is logical. But between scores of 0% and 100%, none of these four feedback types crossed the 0.5 probability threshold. In other words, EOC scores have a definite effect on feedback, but it on average never doesn't determine whether a student's feedback is positive or negative. Rather, all feedback regarding how much time students had to complete their material is assumed to be negative, and all other feedback is assumed to be positive. Thus, students consistently feel confident in their performance and that the material in these textbooks is useful and valuable. But, they consistently feel that they do not have enough time to absorb said information, regardless of how well they performed.

We examined EOC scores and how they change as the chapters progress. The first linear model created predicted EOC score as a function of both chapter_number and textbook. The adjusted R^2 was 0.7169. This model and a graph of the points showed that the advanced college textbook had the best EOC and the high school textbook had the lowest EOC, but all textbooks exhibited a decrease in EOC as chapters progressed. We then made a linear model predicting EOC score as a function of only chapter_number. It had an adjusted R^2 of 0.8602.

We then saw if pulse question responses varied across chapters. The graph shows the four pulse question results were very similar across all of the chapters. We would expect that as students' EOC results decrease, so would their pulse scores, but this does not happen. This indicates one of two possibilities, the pulse question results are not representative of student ability or the students overestimate their ability in the later chapters. In our academic experience, nongraded questions are not taken as seriously as graded ones.

We were also interested in finding information about student-level data. To do this, we grouped the data by student ID, then created variables based on averages of a selection of variables in the page views, checkpoints EOC, checkpoints pulse, and response datasets. We built a linear regression with EOC as the response variable and another with pulse response as the response. Then, with both of these models we built stepwise regression models. The EOC model had an R^2 of .27, while the pulse response model had an R^2 of only .08.