In Fall 2016, I compared a Labster Virtual Lab on Enzyme Kinetics to an Excel Spreadsheet activity which I had used in previous semesters to teach students about enzyme kinetics. To make sure I was making a fair comparison, I split the class into two groups (i.e. Labster First or LF or Excel-First or EF) alphabetically by first name. Percentage of correct responses and Fold improvement (relative to the class averaged initial performance on the assessments) was tracked using different sets of questions given to students at 3 points during the semester as online quizzes using Safe Exam Browser and my campus’ LMS. Assessment items were grouped according to either their source or according to whether they assessed student learning at a recall level or at an applied level (i.e. ACS Q’s, ACS 2007 Q’s, all Applied Q’s, Recall-only Q’s, or All Q’s). The ACS 2007 items have a difficulty index which allowed me to compare the level of student performance relative to national averages. All students to a pre-test (Pre) prior to doing the enzyme kinetics activities. Then, within about 3 weeks, the Labster-First (LF) group completed the Labster Virtual Lab on Enzyme Kinetics while the Excel-First (EF) group completed the traditional excel spreadsheet assignment and all students then took a Mid-test (Mid) consisting of the same items that were on the pre-test. The students then switch activities and the Labster-first (LF) group then completed the excel assignment and the Excel-first (EF) group completed the Labster Virtual Lab. Finally, all the students completed a Post-test (Post) consisting of the same items that were on the previous assessments/quizzes. Student performance on each assessment item and quiz was tracked and the average and variance of the performance of each student group on each assessment item was tabulated. The percentage of correct answers of the different student groups (LF or EF) on each quiz (Pre, Mid, and Post) was then calculated along with the standard deviation. Since the assessment items have different difficulties, the variance in these reported observables was quite high. A “fold-improvement” observable was also calculated based on the subsequent performance of a given group on a particular set of assessment items relative to the initial pre-test performance value for the entire class on that item or items. Standardized performance of students answering selected questions from the ACS 2007 Biochemistry exam or GOB exam is given as a benchmark (i.e. “ACS Benchmark”, blue bars below) for the final level of student achievement in the course. Significant differences between student group achievement at given timepoints are indicated by the probability (p-value) that that result might have been obtained by chance (i.e. the NULL hypothesis) using a 2-tailed Student’s t-test. Pre-test (red), mid-test (yellow), and post-test (green) data for the LF group (solid columns) and EF (transparent columns) are shown for different sub-sets of test items/questions below. The number of questions/items is given beneath each pair of graphs for that set of questions/items.
The data above indicates a significant improvement in student achievement on ACS items. The LF student group had access to the Labster Virtual Labs for a longer period of time than the EF student group and this is correlated with a significantly greater fold-improvement on the post-test (p=.03).
N=3 Q’s

Using a smaller subset of test items from the ACS 2007 Biochemistry Exam, the trend in the results is the same. Although the p-value of the final post-test comparison between the LF and EF groups is a bit higher (p=.14), the more immediate assessment of the effectiveness of the Labster Virtual Lab versus the Excel assignment still indicates a significant improvement in student performance in the Labster-group whereas the Excel assignment did not seem to improve student achievement on the three items used for assessment. It should be noted that ALL items from the ACS 2007 Exam that were even remotely relevant to enzyme kinetics and which had a good discrimination index were used for assessment (i.e. there was very little selectivity in terms of what items were used for the assessment since ACS has already done a very good job at identifying excellent assessment tools for the learning objectives of this course).
The data for questions/items from either the ACS or relevant textbook question banks that assessed students at an applied level is given above. The trend is similar to that found for the ACS questions which typically assess students at an applied level. The criteria for what constitutes an “applied” question is anything which asks the student to apply a concept we have learned about in lecture to a new setting rather than restating a fact or definition (which I defined as a “recall” type question).
There were many more questions in the quizzes which assessed student learning at a recall level. Most of these questions were from the question banks that came with the textbook I used for my course. There was clearly a trend towards increased student learning as the semester progressed. However, there was a small but significant decrease in achievement amongst students in the LF group relative to the EF group. I attribute this to the fact that while the Labster Virtual Labs do a fine job at teaching applied skills, it comes at the cost of reducing the amount of time students have to study/memorize facts which are assessed using traditional recall-type questions. I believe applied skills are much more helpful to students in the long run but I continue to use many recall-type questions in my assessments since many students are used to being assessed in this manner and will complain if assessments become too heavily focused on applied skills.
Comparing student performance on all question types yields a similar trend to that observed for the recall-type questions. 13 questions out the 44 total which were assessed were deemed to not have sufficient discriminatory power (either because they were far too easy or far too difficult) to be useful. These were not included in the other subsets of questions discussed above.