

COMPARING INSTRUCTOR CREATED VS. EXTERNALLY CREATED HOMEWORK ASSIGNMENTS AND THEIR EFFECTS ON EXAM SCORES

David Swart, Lynette Hudiburgh, and Karsten Maurer
Department of Statistics
Miami University
swartdc@miamioh.edu

Many instructors assign homework for their classes through external publishers due to the convenience of having an online platform to deliver and grade assignments for hundreds of students. With advances in technology, however, learning management systems, such as Blackboard, Canvas, and Moodle can be used to build, assign, and grade homework almost as easily as those externally-created programs. Now online instructor-created homework assignments can compete with externally-created assignments in terms of convenience, but can they also compete in terms of quality and student learning outcomes? This study carried out in an introductory statistics class at a Midwestern university will be analyzed to see whether instructor-created assignments could be superior to these externally created assignments in student learning and engagement.

INTRODUCTION

Textbook publishers often provide instructors with user-interfaces through which they can administer homework, quizzes, and other learning aids. These programs can be powerful, capable of randomly generating seemingly endless versions of a given problem set with unique numbers and solutions, and providing a variety of support resources should a student struggle. However, there is one common complaint heard from students in the introductory statistics program at our school; the exams looked different from the homework. The wording in the homework assignments and some of the contexts they focus on differ from our exams, so students have trouble realizing they are looking at the same overarching material. This led to our motivating questions: (1) Could we build and implement homework assignments that were more consistent with our exams? (2) Would those homework assignments lead to better exam performance and student engagement with the material?

The study was carried out in the introductory statistics class for non-statistics majors at Miami University, a mid-sized public university in the midwest United States. Each week, students are required to watch several lecture videos created by the course coordinator. These lectures go along with an instructor-created course pack that includes fill-in-the-blank styled notes, practice exercises, and lab activities. The notes follow the video lectures, which are not tied to any particular textbook. After watching the videos, students take a self-assessment quiz through Canvas, our learner management software, to determine their level of understanding. Students are expected to complete these lectures and notes early in the week to prepare for a large group lecture and problem session. Later in the week, each class then breaks into smaller lab sections to work through an activity which applies a portion of the material. By Friday evening, students are expected to complete a homework assignment and quiz through Pearson's MyStatLab (MSL) software that corresponds to an e-textbook provided through MSL. The problems in the MSL homeworks are all selected from a question bank in MSL, corresponding to a given chapter of the text. In the Fall 2017 semester, we began building our own homework to be more structurally consistent with the rest of the course.

PILOT STUDY

The curriculum in each semester is divided into three four-week units with an exam at the end of each. The first week serves as an introduction to the program, while the last week is dedicated to preparing for finals. We conducted our pilot study during the middle unit using the second exam scores as our response variable of interest. During this second unit, we administered two homework treatments: publisher-created homework (PCH) assignments from MSL, and instructor-created homework (ICH) assignments in Canvas. The rest of the course materials and activities would remain the same. When structuring our treatment groups, we had to consider the

time of the class (six sections) and the instructor (two instructors) as potential effects, so we randomized within each section. In total, 93 students across the six sections volunteered to participate in this initial testing phase. Half of the volunteers in each section were assigned to each homework set as the study began. Ultimately, our results were inconclusive, largely due to a shrinking sample size from attrition of our volunteers, but we are able to apply what we learned to help us prepare for the next phase of the study.

Quantitative Results of the Pilot Study

Using the results of the second exam, we carried out a two-sample t-test to determine whether the average exam score was significantly different between each homework group. Our original intent was to analyze the effects associated with instructor and section, as well as whether the homework and exam scores from the first four weeks had a significant effect on exam 2 scores. However, due to sample size decay, we no longer had enough subjects to estimate the instructor and section effects. Therefore, we tested the mean exam score of the remaining thirteen students in the ICH group against the mean exam score of the remaining forty-four students in the PCH group. A 95% confidence interval for the mean exam score of the ICH group minus the mean exam score of the PCH group is (-6.53, 22.3).

Qualitative Results of the Pilot Study

A survey consisting of twelve questions designed to check for understanding of the content, interest in the homework assignments, and student engagement with the material was administered, along with the second exam. A Likert scale with five levels, from strongly disagree (-2) to strongly agree (2), was used to score each student’s responses to survey items. Several questions were reverse coded to account for students only selecting one response for every item and to preserve internal consistency (Cronbach, 1951). We intended to average the results from these surveys into an overall score measuring three desired attributes: perceived preparedness, engagement with the material, and perceive educational value added. However, Cronbach’s alpha scores of 0.49, 0.23, and 0.36, respectively, revealed our items to be inconsistent. Since alpha is an average of correlations between items on the survey, it is bounded between 0 and 1 with results close to 1 being desirable (Cronbach, 1951). One issue we had was confounding interest, a sign of intrinsic motivation, with extrinsic engagement, which could explain some of the inconsistencies. As a result, we decided to focus on three key individual survey items: “I understood the material on the exam”, “I found the homework assignments engaging”, and “I found the homework assignments interesting.” We will refer to these items as Understanding, Engagement, and Interest, respectively. Figure 1 shows the percentage of students giving each response, faceted by each of the three items. In general, the students responded more negatively to the ICH than the PCH. We attribute these results, in part, to the students being exposed to MyStatLab for four weeks before the pilot study, and the fact that our early instructor-created assignments were not user-friendly.

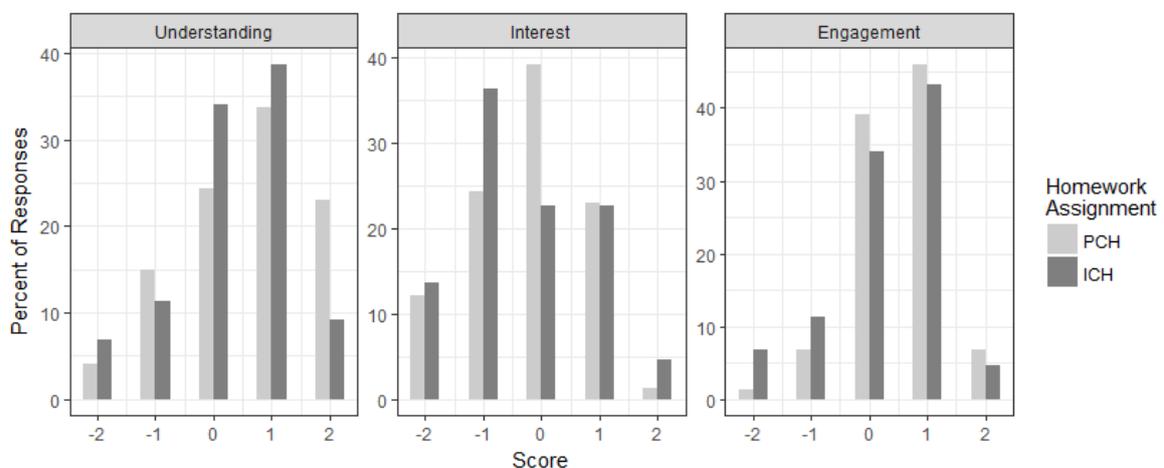


Figure 1: Plot of survey results for questions on understanding, interest, and engagement.

THE CURRENT STUDY

For the Spring 2018 semester, we redesigned the structure of the study to allow better comparison of the instructor-created (ICH) and publisher-created (PCH) homework assignments. First, we have decided to randomize on the section level rather than within the sections. To maintain consistency within each section, each student in a section would be assigned the same homework treatment. Second, we are assigning the homeworks and gathering data for the entire semester, rather than a single unit. Finally, we asked students for permission to use their grades rather than asking them to volunteer to participate in a randomized treatment. We believe this will alleviate the attrition problem from the pilot study, as students will be more likely to grant us permission to use their anonymized grade data for educational research. An instructor effect could be of concern. With only two instructors teaching six sections of the course, one instructor would be assigned two sections of the ICH treatment and one PCH treatment, while the other instructor would be assigned the opposite combination. The time of each section could also be an issue due to the enrollment schedule by student year and honors status, so we control for this using a blocking structure. Given prior data, the early morning and late afternoon sections tend to perform more poorly than the other sections due to the extreme times; therefore, we did not want both of those classes to have the same treatment. The other four sections tend to be roughly homogenous in terms of academic performance, so we instituted a time block separating these inner classes from the two outer classes. Within this inner block, all instructor-treatment combinations were represented.

We have learned much more about working with Canvas and have taken many steps to make our ICH as clear and manageable as possible, while maintaining our desired rigor. Instead of giving the students one large assignment for each module that they have to complete in one sitting, we broke each module down into learning objectives, and created smaller, more manageable assignments for each objective. Canvas has a variety of different question/answer types available, including multiple choice, and dropdowns, so we changed the types of question/answer types used for some questions to give students more control over their answers. One issue, for example, is that dropdown questions require the instructor to enter all acceptable answers, even if the answer is numerical. If the answer to a question was 7, students would be marked incorrect if they answered 7.0, or 7.00; and, if we asked them to round, then things got even more messy. Luckily, we found a way to fix this by using question structures designed to collect numeric response, which allowed us to set a margin of error for the answer.

Since this study is still ongoing, results are forthcoming. Our plan is to average the first and second exam scores as a response for exam performance. We will also administer a reworked version of the survey from the pilot study, using more redundant questions as suggested in Cronbach's paper, so our future results should be more internally consistent and accurately measure student engagement, interest, and perceived understanding. (Cronbach, 1951).

Methods

After the Spring 2018 results are collected, we intend to use the following linear model for analyzing the data. Our response, exam performance for the first two exams, could be modeled as

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \delta_k + \varepsilon_{ijkl}$$

where $\varepsilon_{ijkl} \sim N(0, \sigma^2)$, $\delta_k \sim N(0, \sigma_\beta^2)$

where Y_{ijk} is the exam performance for the l 'th student in the j 'th instructor in the k 'th time block for the i 'th homework assignments group. μ is the overall mean exam performance across all sections, α_i is a fixed effect associated with the i 'th treatment group, β_j is a fixed block associated with the j 'th time block, δ_k is a random effect associated with the k 'th instructor, and ε_{ijkl} is the random error term. $i = 0$ or 1 , $j = 0$ or 1 , $k = 0$ or 1 , and $l = 0, 1, \dots, n_{ijk}$, where n_{ijk} is the number of students in the section for the k 'th instructor in the j 'th time block and i 'th treatment group. More information on this type of modeling can be found Montgomery's *Design and Analysis of Experiments* (Montgomery, 2012).

With all possible instructor-treatment combinations represented in the inner time block and the addition of the outer time block, the treatment and instructor effects will be estimable, as well as the time effect. (Montgomery, 2012). While the grade data and analysis may be the most

important results, it could still be worthwhile to carry out an analysis of variance to see from where, exactly, the variation in student grades is coming, and if the treatment effect from our new homework assignments is making a significant contribution to that variation.

CONCLUSION

Currently we have nothing conclusive to report. The instructor-created assignments have come with their own share of difficulties due to input requirements for some of the question types, however, as the semester progresses constructing them has gotten easier. We are fortunate to have the assignments from the pilot study as a baseline for the second unit, in particular, though as previously stated, significant changes have been implemented to make them more user-friendly for the students. We look forward to sharing our findings from the active study this Summer at the International Conference on Teaching Statistics.

ACKNOWLEDGEMENTS

We would like to thank Dr. Byran Smucker of Miami University for his suggestions on the experimental design.

REFERENCES

- Agresti, A., & Franklin, C. (2016). *Statistics: The Art and Science of Learning from Data (Custom Edition for Miami University)* (4th edition). London: Pearson Education.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.
- Montgomery, D. (2012). *Design and Analysis of Experiments* (8th edition). Hoboken: John Wiley and Sons Inc..