Studying Assignment Size and Student Performance Using Propensity Score Matching

Shirin Mojarad
McGraw-Hill Education
281 Summer Street,
Boston, MA USA
Shirin.mojarad@mheducation.com

ABSTRACT

Teachers and instructors assign students homework of varying lengths. There is considerable evidence that factors such as cognitive load play a role in student performance and learning, but there has not been sufficient study of how these phenomena play out in the specific case of the length of homework. In this paper, we study the impact of assignment size on student performance. This paper represents the first attempt we are aware of to study how long assignments should be, in real-world data, in order to maximize student performance and learning. However, natural assignments of different lengths often vary in other ways. We control for this limitation using propensity score matching (PSM), an approach that helps to control for variables affecting outcome besides the intervention of interest. As such, we can conduct our analysis on large-scale data naturalistically collected through a digital educational platform. We use PSM to study the effect of assignment size on student performance while controlling for assignment difficulty, discrimination and reliability. We find that shorter assignments result in higher performance. These results can be used as a guideline for instructors and instructional designers when designing course assignments.

Keywords

Propensity score matching, assignment size, classical item analysis, item difficulty, item discrimination, student performance, test reliability

1. INTRODUCTION

Graded assignments are used as an effective method to improve students' performance on final tests and improve learning [1]. Considering multiple shorter assignments as opposed to few, larger assignments is amongst the recommendations by USC for designing effective homework assignments [2]. This is because shorter assignments are less intimidating and help enhancing student motivation by minimizing the negative effects of a poor grade on student learning experience. In this study we investigate whether assignment size affects student performance. Since assignments of different sizes often vary in other ways, other assignment characteristics affecting the performance should be isolated to enable the study of assignment size effect on student performance.

Randomized control trials are considered the gold standard in conducting studies to investigate the effect of a particular intervention on a specific outcome [3]. However, their application is limited in educational settings as they can be conducted on a limited number of students. Results from the comparison of RCTs and OSs show that OSs can expand upon RCTs due to the use of large and diverse sample population [4]. Propensity score matching (PSM) is a common method in OSs to study the causal effect of an intervention on a particular outcome [4]. In this paper, we have used PSM to leverage the large amounts of data available through McGraw Hill Education digital platforms.

The goal of this paper is to study the impact of assignment size on student performance in isolation from other assignment characteristics including assignment difficulty, discrimination and reliability. This is the first effort of its kind in measuring an optimal assignment size to maximize student performance.

2. Materials and Methods
2.1 Data

We study these issues using data from assignments completed through McGraw-Hill Education's higher education platform, Connect. Connect is one of the most widely used digital platforms in higher education with over two million students and 25,000 instructors [5][6]. Connect allows instructors to design assignments in form of homework, practice, exams, or quizzes. Here, we refer to assignments as a set of items that either test student on knowledge and skills or allow students to practice what they have learnt on the course. Most Connect assignments are graded by the system automatically.

The dataset in this study is retrieved from all the courses created from the title Managerial Accounting 2nd Edition, by Robert Libby. We include all data for this title between September 2014 and January 2016. The original dataset included 362 classes, where 12,588 students responded on 3,072 items on 5,330 assignments, for a total of 1,031,298 student-item pairs. We have kept only assignments that have 10 or more student submissions. After applying this filter, there are 2,826 assignments left in the data. From the four of types of assignments in Connect, i.e. homework, practice, exam and quiz, we have focused on homework assignments. The reason is that in homework assignments, score is not as a strong motivator as exams and quizzes since homework assignments have a low weight in final score and are mainly aimed for development of self-study habits in students [7]. Hence, students are more motivated by learning to finish homework assignments. Therefore, size of a homework assignment will be an important factor in keeping students engaged throughout the assignment.

2.2 Exploratory Data Analysis

The dataset includes assignments' size, difficulty, discrimination, reliability, and average score where difficulty, discrimination and reliability are calculated using classical item analysis [8]. The assignment size in this dataset varies between 1 to 101 items.

Based on the rarity of very large assignments (and the likelihood that an assignment with over 100 items represents test practice or something different than briefer assignments), we have filtered down to assignments of size 16 or less. Filtering in this fashion still retains 98% of the assignments. We categorized assignments into short and long assignments by using a cut off for number of items within that assignment. Frequency of assignments drops for assignment sizes of larger than 5, which indicates most instructors prefer shorter assignments of size 5 or less. We have used this as a reference to decide a cut off value for number of items for short and long assignments. Following this definition, there were 1,787 short and 1,039 long assignments.

Figure 4 shows the mean score of different assignment sizes. As shown in this figure, the mean score of assignments drops as the assignment size increases.
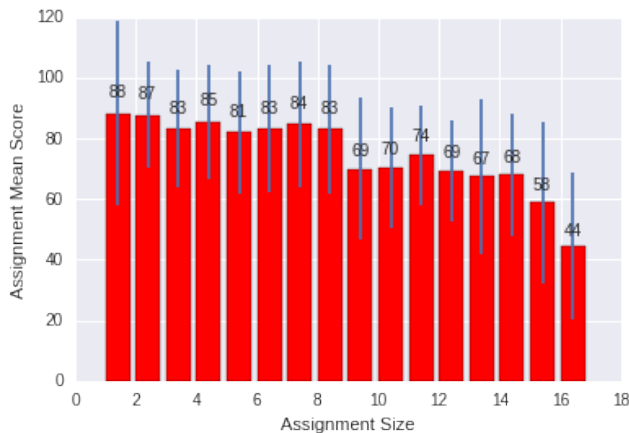

Figure 4. Assignment size versus assignment mean score

3. Results

Overall, students achieve an average 8.7 (on a scale of 0 to 100) higher score on short assignments than long assignments. When we control for difficulty, discrimination and reliability using PSM, students still achieve a 6.8 (on a scale of 0 to 100) higher average score on short assignments compared to long assignments.

The differences between the characteristics of short and long assignments matched using PSM are shown in Table 2. We have used Algina's d to compute the effect size of the difference of means between the two assignment groups [9].

As shown in this table, the effect size of difficulty, discrimination and reliability between two groups of assignments is negligible, indicating that these factors are no longer significant once we control for them using propensity score matching.

Table 2. P-value and the effect size of short versus long assignments, matched using PSM method

Attribute	Mean Difference	Effect Size (Algina's d)	P-value
Average Score	6.8	0.40	<0.001
Difficulty	0.00	0.00	0.99
Discrimination	0.00	0.01	0.53
Reliability	0.00	- 0.01	0.44

4. Conclusion

In this study, we investigated the effect of assignment size on student performance. Results of EDA show that student performance drops as the assignment size increases. The relation between assignment size and average score indicated that performance drops dramatically in assignments sizes of higher than 6. Hence, we used a cut off value of 6 to define short and long assignments. In order to investigate the statistical significance of this difference in two groups of assignments, in isolation from other factors affecting assignment performance, we used propensity score matching (PSM). The effect size and average performance difference of short versus long assignments is still significant when matching assignments with similar difficulty, discrimination and reliability. This indicates that longer assignments may increase cognitive load for students and negatively affect student performance and learning. These results can be used in form of recommendations to instructors when they are designing homework assignments on the Connect platform.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Latif, E., Miles, S., (2011). The impact of assignments on academic performance. *Journal of Economic and Economic Education Research*, (Nov. 2011), 12 (3).

[2] Center for Excellence and Teaching, University of Southern California,http://cet.usc.edu/resources/teaching_learning/docs/teaching_nuggets_docs/4.2_Assignments_and_Homework.pdf, last accessed at March 2016.

[3] Silverman, S. L. (2009), From Randomized Controlled Trials to Observational Studies, *The American Journal of Medicine*, Volume 122, Issue 2, Pages 114–120.

[4] Rosenbaum, P. R., and Rubin, D. B., (1983). The Central role of the propensity score in observational studies for causal effects. *Biometrika*, 70 (1), 41-55.

[5] Feild, J., 2015. Improving student performance using nudges analytics. *Educational Data Mining*.

[6] Agnihotri, L., Aghababyan, A., Mojarad, S., Riedesel, M. and Essa, A., (2015). Mining Login Data For Actionable Student Insight. In Proc. 8th International Conference on Educational Data Mining.

[7] Sharma, Y. K., Fundamental Aspects of Educational Technology. Kanishka Publishers, Distributors New Delhi.

[8] Smith, Jeffrey K.. (1987). Review of *Introduction to Classical and Modern Test Theory*. *Journal of Educational Measurement* 24 (4). [National Council on Measurement in Education, Wiley]: 371–74.

[9] Algina, J., Keselman, H. J., & Penfield, R. D. (2005). An alternative to Cohen's standardized mean difference effect size: A robust parameter and confidence interval in the two independent groups case. *Psychological Methods*, 10, 317–328.