Assessment ~~of~~ for Learning
Studying Assignment Size and Student Performance Using Propensity Score Matching

Shirin Mojarad, McGraw-Hill Education

Research Background

"Most students become similar with regards to learning ability, rate of learning and motivation for further learning when provided with favorable learning conditions" – (Bloom, 1978) [1]

A series of well-designed formative assessments is an essential part of a mastery learning

- To ensure students have required knowledge and skills for new learning tasks.
- To increase students' confidence in learning new tasks, knowing they mastered the previous one.
- To improve students' learning effectiveness and motivation by providing regular feedback

Research Question

Can we design better formative assessments to improve students' performance through appropriate and timely feedback?

How assignment size affects student performance

- Assignment is a formative assessment
- Assignment size is defined as the number of items it entails

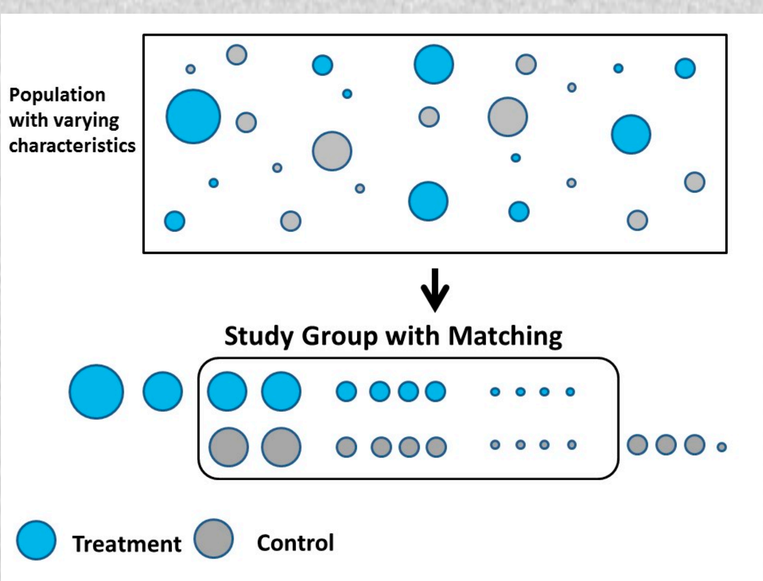
Data

- Connect Course on Managerial Accounting [2,3]
- 362 classes, 5,330 assignments
- 12,588 students, 3,072 items
- 1,031,298 student-item pairs

Methodology

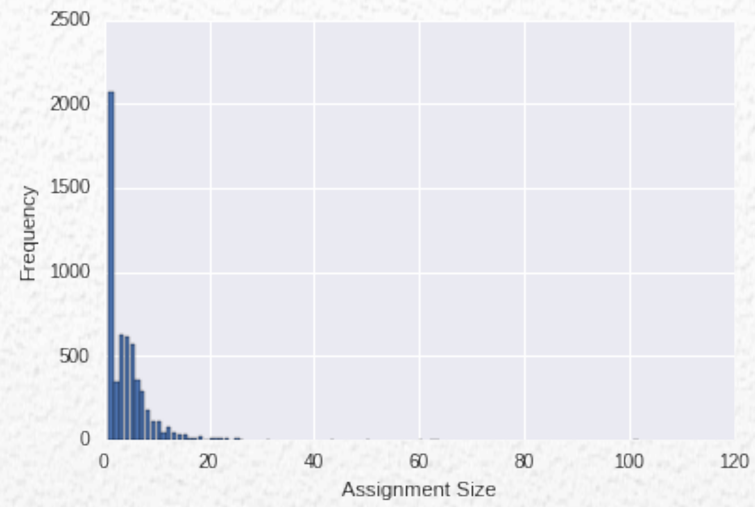
- Define and label short and long assignments
- For each short assignment (treatment group), identify a long assignment (control group) that is as similar as possible in difficulty, discrimination and reliability and compare performance [4]
- Average the difference in performance between the matched pairs

Propensity Score Matching (PSM)

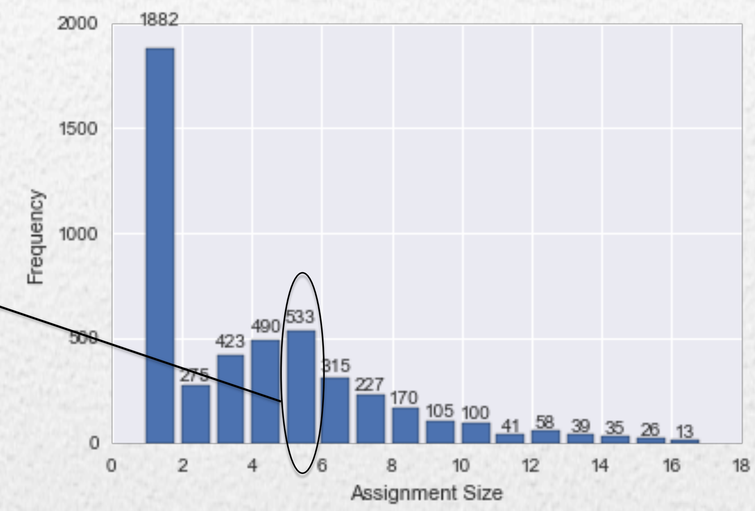


Exploratory Data Analysis

98% of activities have 16 items or less

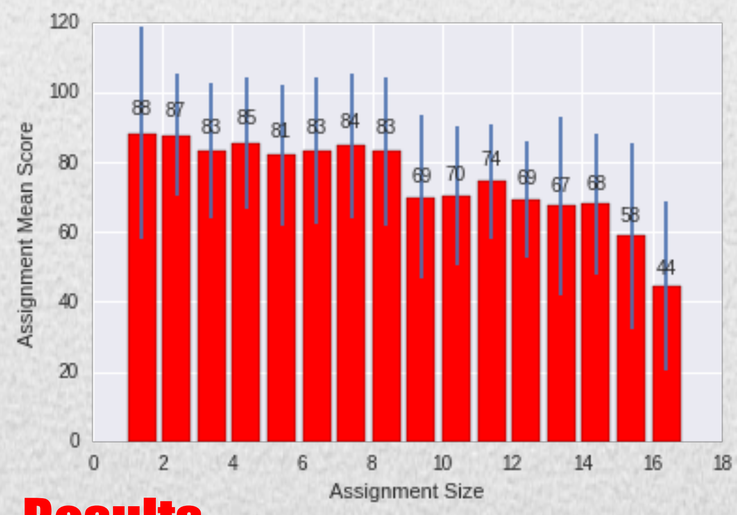


Most instructors prefer shorter assignments of size 5 or less.



Following this definition, there were 1,787 short and 1,039 long assignments.

Average assignment performance drops as assignment size increases



Results

Before Matching: average 11.1 (scale of 0-100) higher score on short assignments than long assignments

After Matching: average 6.8 higher score on short assignments than long assignments

Attribute	Median Difference	Effect Size (Algina's d [5]	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

[1] Bloom, B.S. (1978). New views of the learner: Implications for instruction and curriculum. *Educational Leadership*, 35(7), 563-676.
[2] Feild, J., 2015. Improving student performance using nudges analytics. Educational Data Mining.
[3] 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.
[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] 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.