Embedded Tutoring and Successful Course Completion (2019)

Background:

During the spring 2019 and fall 2019 semesters, roughly 140 sections across 20 disciplines had embedded tutors in the classroom. The embedded tutors were also available to help students outside of the classroom. One of the primary goals of this intervention was to increase successful course completion rates (C or better final grades). This research brief examines whether students in sections with embedded tutors have higher success rates than students in similar sections without embedded tutors.

Methodology:

Because neither embedded tutors nor students were randomly assigned to sections, one of the fundamental assumptions behind the standard counterfactual framework may not be satisfied. If group assignment is confounded with the outcome of interest, then the average treatment effect associated with exposure to an embedded tutor cannot be directly determined without attempting to balance the groups on the observable covariates (Guo and Fraser 2014).

While there are several methods available to accomplish this task, I used coarsened exact matching (King, Blackwell et al. 2010, King, Nielsen et al. 2011, Iacus, King et al. 2012, King and Nielsen 2015) primarily because this method provided for exact matching on the covariates available. Students in the treatment group (enrolled in a section with an embedded tutor) were matched with control students (enrolled in a section without an embedded tutor) on several observed covariates that correlate with student success. Treatment and Control groups were matched exactly on the following variables using data from the same and previous terms:

Course Level Variables:

- Course
- Instructor
- Campus Location
- Time of day
- Number of weeks
- Section Limit (Max enrollment)

Student Level Variables:

- Prior GPA (range)
- Gender
- Underrepresented Ethnic Minority Status
- Received BOGW
- Academically Disadvantaged Status
- Full-time Status
- Math Placement Level

- English Placement Level
- Age (over/under 24)

After matching on the aforementioned observed covariates, the Average Treatment Effect (ATE) was assessed. Finally, because students are nested within sections, the intraclass correlation coefficient (ICC) was calculated to assess the percent of variance that exists between groups (sections).

Results:

Two thousand treated students, out of 2038, had exact matches on the course and student levels variables cited above. Because the treated were matched exactly, the degree of imbalance, as measured by the L1 statistic, was zero. The difference in success rates between the treated students and matched control students was not statistically significant (p=.799) (See Table 1 below).

Table 1.

```
Treatment-effects estimation                   Number of obs      =      4,003
Estimator      : nearest-neighbor matching     Matches: requested =          1
Outcome model  : matching                                     min =          1
Distance metric: Mahalanobis                                  max =          7
```

success	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATE						
embed (Embedded vs Not Embedded)	-.0037763	.0148262	-0.25	0.799	-.0328352	.0252826

The intraclass correlation coefficient (ICC) was roughly .12, indicating that about 12% of the total variance in success rates was between sections. Given the magnitude of the ICC and the size of clusters, standard errors were corrected to account for the design effect (Raudenbush and Bryk 2002). Using the cluster robust standard errors increased the p-value to .98. In other words, over an infinite number of samples, one would be likely to find the observed differences by chance 98 out of 100 times.

Though an attempt was made to balance the treatment and control groups on the available, observable covariates, there are undoubtedly other unavailable observed and unobserved covariates that may confound the results. To assess the magnitude at which an unobserved covariate could alter the results of the statistical test, Rosenbaum's (2005) sensitivity analysis was performed. The results indicate that an excluded observed or unobserved covariate would need to increase the odds of exposure to embedded tutoring by a factor of 50 to change the results of the statistical test. In other words, the results of this test are extremely robust to potential confounders.

Discussion:

In sum, the average treatment effect of embedded tutoring on successful course completion was not statistically significant. This could be due in part to low student utilization of tutoring opportunities outside of the classroom. To this point, the college will now record each student's tutoring attendance in order to assess whether the lack of an observed effect was due in part to low participation in tutoring sessions. While this does introduce another potential confound (motivation or grit), it may show that similar students who actually participate in tutoring do indeed outperform those who do not.

References:

Guo, S. and M. W. Fraser (2014). Propensity score analysis: Statistical methods and applications, Sage Publications.

Iacus, S. M., et al. (2012). "Causal inference without balance checking: Coarsened exact matching." Political Analysis: 1-24.

King, G., et al. (2010). "cem: Coarsened exact matching in Stata."

King, G. and R. Nielsen (2015). "Why propensity scores should not be used for matching." Copy at http://jmp/1sexgVw Export BibTex Tagged XML Download Paper **481**.

King, G., et al. (2011). "Comparative effectiveness of matching methods for causal inference." Unpublished manuscript **15**.

Raudenbush, S. W. and A. S. Bryk (2002). Hierarchical linear models : applications and data analysis methods. Thousand Oaks, Sage Publications.

Rosenbaum, P. R. (2005). "Heterogeneity and causality: Unit heterogeneity and design sensitivity in observational studies." The American Statistician **59**(2): 147-152.