In-Person Tutoring at SLCC: Does it Work?

Jeff Webb
Office of Institutional Research and Reporting,
Salt Lake Community College

July 28, 2017

Introduction

SLCC has been offering tutoring services since 1986 when the Learning Center Program was created. In July 2017, this program was reorganized as the STEM Learning Resources department within the School of Science, Mathematics, and Engineering. Online tutoring was first offered as an alternative to in-person tutoring in 2012 but was mostly unsuccessful. In January 2017, the college began offering online tutoring through Brainfuse, which students accessed through their Canvas course pages. In the process of researching the impact of Brainfuse use on student course performance, we considered the impact also of SLCC's traditional tutoring options. This report summarizes those findings.

Summary of Findings

In person tutoring appeared to help students get higher grades, on average, than they otherwise would have in the courses for which they sought help. The average grade improvement due to tutoring was .22 on a 4 point scale. Thus, we can say that tutoring helped students improve course grades by approximately one grade increment (the difference, for example, between B- and B).

Research design

Course grade was the outcome variable in the study, and the treatment was tutoring. We were interested in whether tutoring *caused* improved performance in the classes for which students sought help. Our analytic strategy was to compare grade outcomes for students in the treatment condition (tutoring) to those in the control condition (no tutoring). One major threat to the validity of causal inference in observational studies such as this one is selection bias: students who seek tutoring may have performed better with or without tutoring due to unobserved differences in ability, preparation or motivation. One method for countering selection bias is propensity score matching (PSM), which we used in this study. We matched students who used tutoring with students who did not, on a variety of characteristics. PSM largely eliminated systematic differences between the groups, making causal inference—our analytic goal—more plausible.

We focused on the 4 most recent terms (excluding summer): Fall 2015, Spring 2016, Fall 2016, Spring 2017. We required that matched pairs of students had the *same classes* with the *same teachers* (though not necessarily in the same term) and were as similar as possible on the following characteristics:

- cohort year
- cumulative earned credits
- term credits attempted
- cumulative GPA
- age at beginning of term
- gender
- ethnicity
- former concurrent enrollment
- term pell eligibility

- declared transfer

The goal in matching is to create treatment and control groups that, as in a randomized control trial, do not display systematic differences. Students in the control group should be just like those in the treatment group, having opted out of treatment—tutoring—for reasons of chance rather than substance. They are students who *would have used* tutoring if randomness in the world had not intervened. To make this idea plausible we need to show that matching in fact successfully aligned treatment and control groups on the matching characteristics listed above. See Table 1 for a comparison of the matched groups.

Data

Tutoring data was extracted from the TutorTrac system. It contained information on 143160 tutoring visits over 13 terms, from 201240 to 201720:

- Student ID
- Student name
- Date and time of tutoring session
- Originating class
- Term

The data quality was suspect in some cases. We eliminated rows that had missing or obviously incorrect data and calculated a single usage summary—total visits, total time—for each student in each course. (Students could be represented multiple times in the data if they sought tutoring for more than one class.) There were 22430 such usages during this time period by 10235 students for 376 courses.

Data cleaning and modeling

We selected the most recent 4 regular terms (Fall 2015 - Spring 2017, excluding summer) for analysis because tutoring practices, like other services at the college, are likely to be in flux, and data that is older than 2 years may not be representative of current practices. Additionally, the size of the dataset was awkwardly large; filtering based on recency was therefore not only sensible, ensuring representativeness, but also convenient, bringing the dataset down to a reasonable size for statistical analysis.

Students who used Brainfuse tutoring in spring 2017 were removed from the data in order to avoid mistaking the effect of electronic tutoring for that of traditional tutoring. Additionally:

- We construed course grades of W as 0 (equivalent to E). That is why, in Table 6, some students had attempted credits of 0: they withdrew from the course.
- We removed students with missing gender information.

We joined tutoring data to new student cohorts, from Fall 2005 to Spring 2017, as defined by the Voluntary Framework of Accountability (VFA). VFA cohorts offer a consensus way to define new students. Most, but not all, SLCC students are in a VFA cohort. (For example, VFA excludes concurrent enrollment students.) Using VFA cohorts therefore entailed filtering the data still further to a total of 84216 students, 6252 of whom used tutoring.

Matching

Matching was done with replacement (meaning that a control could serve as a match for more than one treatment student) and the matching was 1-to-1 (meaning that there were the same number of students in the control as there were in the treatment condition). Table 1 profiles the matched and unmatched treatment and control groups.

		Unmatched		Matched
tutoring	No	Yes	No	Yes
count	84216	6252	6252	6252
cumulative earned credits	35.23	43.73	43.32	43.73
term credits attempted	10.48	10.48	10.55	10.48
cumulative GPA	2.74	3.04	3.03	3.04
age at beginning of term	24.24	27.06	26.50	27.06
% female	48.78	49.63	49.52	49.63
% white	50.19	45.03	44.58	45.03
% latinx	11.42	15.20	14.65	15.20
% former concurrent enrollment	30.88	22.92	22.62	22.92
% term pell eligibility	37.59	49.36	50.02	49.36
% online	18.36	8.40	10.96	8.40
% declared transfer	24.11	24.18	24.68	24.18
avg grade	2.17	2.33	2.04	2.33

Table 1: Comparison of matched and unmatched treatment and control groups.

In most cases, as can be seen in Table 1, the matching operation made the two groups more comparable. While matching *widened* the difference between the group means for declared transfer and credits attempted, that larger difference was not statistically significant. In all the other cases, as expected, matching brought the groups closer together (this was particularly true in the case of crucial performance-related variables, such as earned credits and cumulative GPA), and the end result, in terms of course performance, was to *increase* the raw performance difference (average course grade of 2.04 for students who did not use tutoring compared to 2.33 for those who did).

Results

We first present descriptive statistics on tutoring usage from the most recent 4 terms of TutorTrac data (fall 2015 - spring 2017).

Descriptive statistics

Table 2 shows tutoring visits by class. Math students were by far the largest users of tutoring. Table 3 shows the classes with the largest non-Math tutoring use.

class	visits
MATH1010	12504
MATH1050	10153
MATH0980	4610
MATH1040	3362
MATH1060	3150
MATH0990	2959
MATH1030	2500
MATH1210	2420
CSIS1410	1448
MATH1220	1238

Table 2: Tutoring visits by class, Fall 2015 - Spring 2017

The average length of a tutoring visit was 30.58 minutes, and the students had, on average, 4.62 visits per course (median visits = 1) and 2.35 total hours of tutoring per course (median hours = 0.7). We did not set

class	visits
CSIS1410	1448
CHEM1110	1128
PHYS2210	1003
MGT2040	850
CSIS1400	790
CHEM1210	737
CHEM1010	417
BIOL1610	390
CHEM1220	295
CHEM2310	280

Table 3: Tutoring visits by class (non-Math), Fall 2015 - Spring 2017

a threshold for what counted as a tutoring visit but included all logged visits, no matter how short. The amount of tutoring used by students varied by course. Tables 4 and 5 show statistics for the 10 heaviest usage courses.

class	max visits	avg visits	median visits	n
CHEM2315	8.00	3.86	4.00	7
ENGR2030	11.00	3.44	3.00	18
MATH2900	11.00	3.78	3.00	9
MATH2280	5.00	3.00	2.50	6
PHYS2020	12.00	3.38	2.50	8
ACCT2600	17.00	3.75	2.00	8
CHEM1120	36.00	4.74	2.00	42
CHEM2310	18.00	3.32	2.00	84
CHEM2320	13.00	3.25	2.00	12
CHEM2325	3.00	1.83	2.00	6
CSIS2420	37.00	4.51	2.00	60
EE2260	3.00	1.67	2.00	9
ENGR2140	3.00	1.86	2.00	7
MATH0980	225.00	6.94	2.00	656
MATH1010	188.00	6.36	2.00	1687
MATH1050	149.00	5.99	2.00	1380
PHYS2010	44.00	5.65	2.00	26
PHYS2220	45.00	5.85	2.00	33
CSIS2810	48.00	9.05	1.75	10
BIOL1620	5.00	2.33	1.50	6

Table 4: Tutoring visits by class, Fall 2015 - Spring 2017. The table is ordered by median visits and includes only classes with more than 5 students.

Figures 1 and 2 show the distribution of visits and time for all students and courses from Fall 2015 to Spring 2017. The large range of the x-axes indicates that a small number of students were very heavy users of tutoring.

Inferential statistics

For statistical analyses we used the matched data, with a total of 12,504 observations split evenly between treatment and control conditions. See Table 1 for summary statistics on the two groups. Inferential results are as follows:

1. Statistical analysis indicated that the raw difference in grade outcomes between the groups was
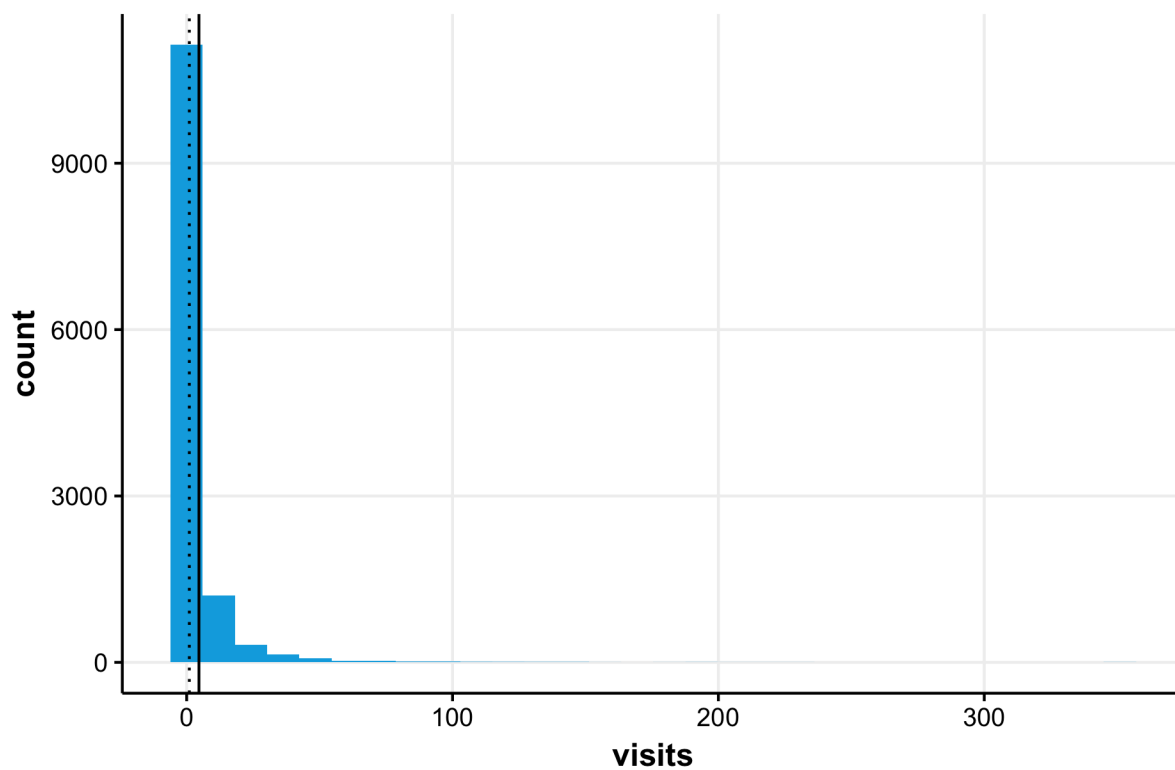
Figure 1: Distribution of tutoring visits per student course, fall 2015 - spring 2017. Mean visits = 4.62 (solid line). Median visits = 1 (dashed line).
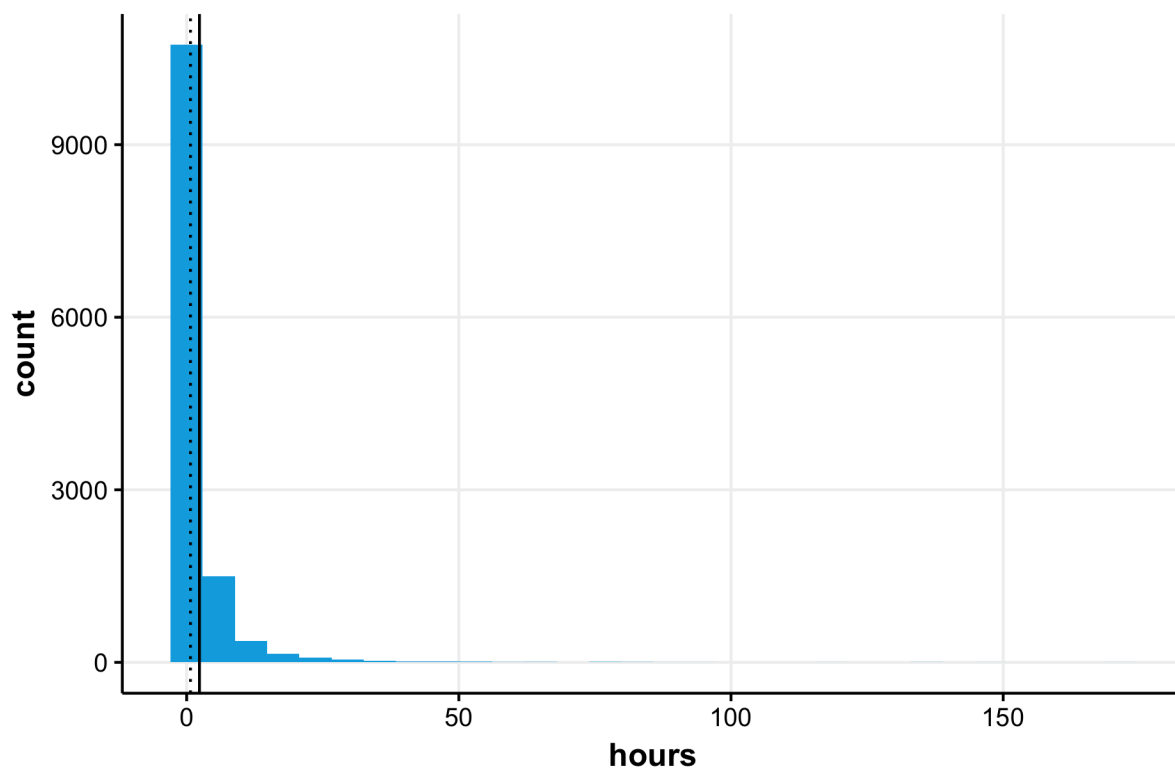
Figure 2: Distribution of total hours spent in tutoring per student course, fall 2015 - spring 2017. Mean hours = 2.35 (solid line). Median hours = .7 (dashed line).

class	max hours	avg hours	median hours	n
CHEM2315	20.03	4.94	2.12	7
EE2260	4.13	1.89	2.08	9
ACCT1120	13.67	3.04	1.86	24
ACCT2010	20.83	2.28	1.44	92
ACCT2020	13.50	2.10	1.42	52
CHEM2325	4.35	1.75	1.35	6
PHYS2220	74.70	5.20	1.30	33
MATH2280	2.70	1.23	1.27	6
BIOL1620	7.83	2.77	1.25	6
ACCT2600	10.33	2.30	1.21	8
ACCT1110	9.92	1.78	1.18	55
ENGR2300	8.05	2.25	1.18	14
MEEN1000	11.92	2.95	1.12	14
MATH1040	81.40	2.79	1.09	702
MATH1010	93.50	2.86	1.05	1687
PHYS2010	27.15	4.10	1.02	26
PHYS2215	16.65	3.15	0.98	13
CHEM1120	34.70	5.14	0.98	42
MATH1030	52.04	2.32	0.97	444
BIOL2325	10.05	1.72	0.97	29

Table 5: Average tutoring hours by class, Fall 2015 - Spring 2017. The table is ordered by median hours and includes only classes with more than 5 students.

statistically significant. Specifically, regression analysis estimates that tutoring improved grades by, on average, 0.22 on a 4 point scale. The 95% confidence interval for this estimate was [0.16, 0.27]. Figure 3 shows the grade distribution by treatment condition.

2. While increased grades were associated with both visits and total time (meaning that as tutoring visits and time spent in tutoring went up, so did grades), the effect was negligibly small.

Figure 4 shows the difference in average course grade, by class, for tutoring vs. no tutoring. The bars represent point estimates, and the error bars represent 95% confidence intervals on those point estimates. As we can see, the confidence intervals are quite wide. While we can have a lot of confidence in the overall grade difference [0.16, 0.27], we don't have a lot of confidence in grade differences by class. In large part this is because of low n: by the time we have subsetted the data to a single class, there are not many observations left, which limits our ability to discern differences. The classes with error bars that do not include 0 are ones in which tutoring had a statistically significant effect on grade outcomes (these tend to be classes with larger n), and overlapping error bars between classes—all of them—mean that we cannot say with confidence that tutoring impacted grades differently in those classes. Thus, we know that tutoring had an overall effect, but we can't confidently rank classes according to tutoring's class-specific effects.

Discussion

There is strong evidence that tutoring helps students get higher grades in the courses for which they sought tutoring help. This conclusion is not simply correlational. Use of PSM allows us to be more assertive in our conclusions: tutoring appears to *cause* higher grades. This is perhaps not surprising inasmuch as SLCC tutors have been trained to offer instruction in the specific curriculum of the courses for which students are seeking help. Still, these results are a nice confirmation of the efforts of SLCC tutors.

One limitation of the study is that causal inference about treatment effects using PSM is restricted to local effects. That is, we can can confidently say that tutoring impacts course grades, but only for the types of
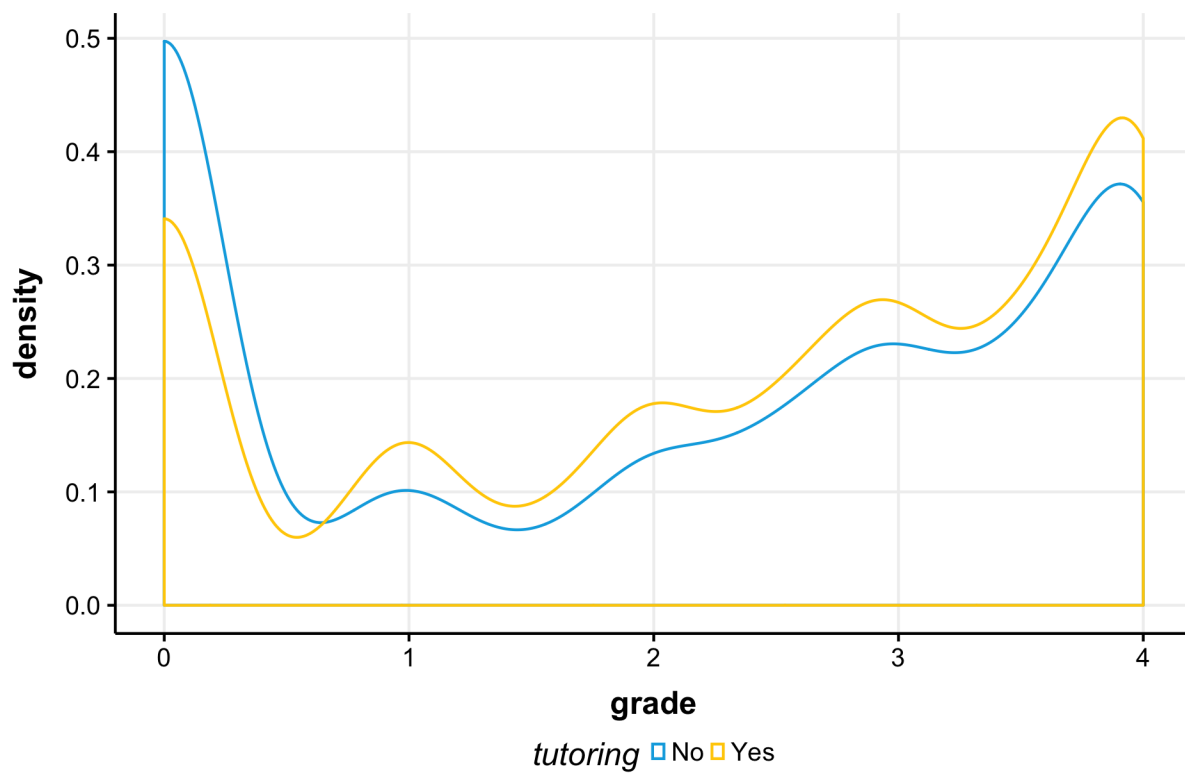
Figure 3: Distribution of course grades by treatment, after matching, fall 2015 - spring 2017. Students who got tutoring (the yellow line) had fewer E and D grades compared to those who did not get tutoring (the blue line) and more grades at the C, B and A levels.
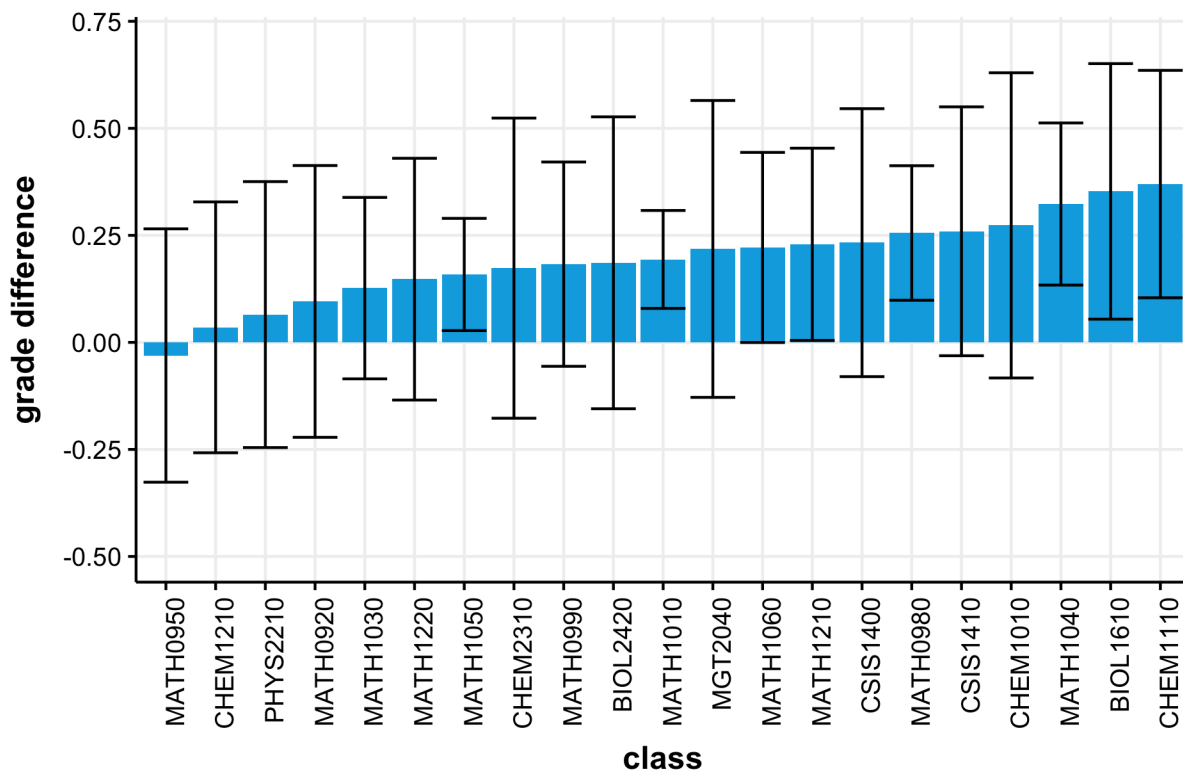
Figure 4: Difference in course grades by treatment and class, after matching, fall 2015 - spring 2017. The error bars represent 95% confidence intervals on the point estimate of the grade difference. The confidence intervals are quite wide. While we can have a lot of confidence in the overall grade difference, we don't have a lot of confidence in grade differences by class. In large part this is because of low n. The classes with error bars that do not include 0 are ones in which tutoring had a statistically significant effect on grade outcomes (these tend to be classes with larger n), and overlapping error bars between classes—all of them—mean that we cannot say with confidence that tutoring impacted grades differently across classes.

students in the specific courses in the matched samples. Table 6 presents the ranges of the continuous student demographic variables in the matched samples. The ranges suggest that the results are widely applicable: they capture the vast majority of students and are comparable between the two groups. An additional limitation is that the matches involving first year students are likely inexact because new students have no cumulative GPA and cumulative GPA is a crucial control on differences in unobservable characteristics like motivation and preparation. There is the possibility of selection bias for these students. Could this selection bias be driving our results? No. When we eliminated the first semester students for the comparison, the grade effect of tutoring went down only slightly.

Future studies could consider retention also as an outcome, which would require using more than two years of data. Future studies might also put a threshold on what counts as a legitimate tutoring session. For the present study, no thresholds were used; all visits no matter how short were included in the analysis. Using such a threshold would likely do a better job of capturing meaningful interactions with tutors.

tutoring	No	Yes
minimum age	14	15
maximum age	78	73
minimum earned credits	0	0
maximum earned credits	247	197
minimum attempted credits	0	0
maximum attempted credits	30	22
minimum cumulative GPA	0	0
maximum cumulative GPA	4	4

Table 6: Minimum and maximum of continuous student demographic variables in matched samples.