UNITED STATES MILITARY ACADEMY

WEST POINT, NEW YORK

HONORS THESIS

Measuring the Effect of Writing Center Visits on Student Performance

by

Joseph Nicholas Zuccarelli

May 2021

Thesis Advisor: LTC Kevin Cummiskey
Second Reader: LTC Andrew Lee

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** May 2021	**3. REPORT TYPE AND DATES COVERED** Senior Thesis (Honors)

4. TITLE AND SUBTITLE Measuring the Effect of Writing Center Visits on Student Performance	**5. FUNDING NUMBERS**
6. AUTHOR(S) CDT Joseph Zuccarelli, 2LT Nicholas Cunningham, LTC Kevin Cummiskey, LTC Andrew Lee, Dr. Colleen Eils	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) United States Military Academy West Point, NY 10996-1786	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) Mounger Writing Center West Point, NY, 10996-1786	**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Distribution Statement (mix case letters)	**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

Many colleges have invested in writing centers to lead institutional efforts to improve student writing across their curriculum. These writing centers often offer one-on-one consultation services in which students receive specialized coaching from trained staff on upcoming writing assignments. However, there are currently few studies in writing pedagogy literature that quantify the effect of these consultation services on future academic performance. In this study, we matched 1000 students who visited the West Point Writing Center from January - December 2019 for appointments with trained staff members before a writing assignment to controls in the student body. We used statistical methods rooted in causal inference (propensity score matching) to select controls with similar demographics, experience, and academic ability. We found a statistically significant improvement in student performance on the writing assignment following multiple visits to the writing center after adjusting for the student characteristics mentioned above. Ultimately, these findings suggest that wider access to consultation services may be beneficial to developing undergraduate writing.

14. SUBJECT TERMS			**15. NUMBER OF PAGES**
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UL

THIS PAGE INTENTIONALLY LEFT BLANK

Measuring the Effect of Writing Center Visits on Student Performance

Joseph N. Zuccarelli
CDT, USMA
B.S., United States Military Academy, 2021

Submitted in partial fulfillment of the
requirements for the degree of
BACHELOR OF SCIENCE
in **Mathematical Sciences**
with Honors
from the
UNITED STATES MILITARY ACADEMY
May 2021

Author: Joseph Nicholas Zuccarelli

Approved by:

LTC Kevin Cummiskey
Thesis Advisor

LTC Andrew Lee
Co-Advisor

Dr. Colleen Eils
Co-Advisor

Colonel Tina R. Hartley
Head, Department of Mathematical Sciences

THIS PAGE INTENTIONALLY LEFT BLANK

United States Military Academy

CDT Joseph Zuccarelli, 2LT Nicholas Cunningham, LTC Kevin Cummiskey, LTC Andrew Lee, Dr. Colleen Eils

7 May 2021

Measuring the Effect of Writing Center Visits on Student Performance

Abstract

Many colleges have invested in writing centers to lead institutional efforts to improve student writing across their curriculum. These writing centers often offer one-on-one consultation services in which students receive specialized coaching from trained staff on upcoming writing assignments. However, there are currently few studies in writing pedagogy literature that quantify the effect of these consultation services on future academic performance. In this study, we matched 1000 students who visited the West Point Writing Center from January - December 2019 for appointments with trained staff members before a writing assignment to controls in the student body. We used statistical methods rooted in causal inference (propensity score matching) to select controls with similar demographics, experience, and academic ability. We found a statistically significant improvement in student performance on the writing assignment following multiple visits to the writing center after adjusting for the student characteristics mentioned above. Ultimately, these findings suggest that wider access to consultation services may be beneficial to developing undergraduate writing.

Acknowledgements

I want to express my thanks to all the individuals who aided me in carrying out this year-long research project. In particular, I would like to thank my three advisors–LTC Kevin Cummiskey, LTC Andrew Lee, and Dr. Colleen Eils. All three of these individuals went above and beyond to ensure that I benefited as much as possible from this experience. I sincerely enjoyed meeting with them each week. I consider them as much more than just advisors–they are mentors for whom I have a great deal of respect and admiration. I also would like to thank my family and friends for their unwavering love and support during both the highs and the lows of this journey. Finally, I would like to thank the USMA Math Department and the Writing Fellows Program for making my cadet experience both rewarding and enjoyable.

Table of Contents

List of Figures

List of Tables

1 Introduction

The purpose of this study is to measure the effect of visiting a writing center on student grades for major writing assignments. Our study focuses on the Mounger Writing Center (MWC) at the United States Military Academy (USMA). The MWC's mission is to engage clients in productive conversations about their ideas and how to express them more clearly, forcefully, and effectively. This mission is achieved through hour-long sessions during which clients receive oral and written feedback from both peer and professional tutors. These tutors, known in the MWC as *Writing Fellows*, are high-performing students who have completed at least three credit-hours worth of coursework in writing pedagogy as well as contracted postgraduates with prior writing center or classroom experience. The overwhelming majority of appointments that clients schedule with these writing fellows are for what USMA refers to as a Signature Writing Event (SWE). SWEs are large, heavily-weighted writing assignments that are typically due at the conclusion of core curriculum courses, allowing time for students to develop as thinkers and writers in the discipline at hand. There is little research on the effect of writing center visits on graded assignments such as SWEs and what examples we may find lack scientific rigor. Existing studies draw conclusions based on correlational evidence derived from simple probabilistic models. Therefore, our study seeks to address this lack of scientific rigor by applying an advanced statistical method rooted in causal inference literature–propensity score matching.

1.1 Writing Center Studies

In its current state, writing center research is lacking advanced quantitative analyses. Although there are a few published quantitative studies within the field, they often feature the application of relatively simple statistical methods such as basic correlation and regression analysis. Although such methods can be used to carry out meaningful statistical analysis, researchers can sometimes use them to draw excessive conclusions that do not follow. For instance, Salem (2016) investigates factors associated with writing center attendance using data from 4,202 students at Temple University. Salem analyzes several potential factors associated with writing center attendance such as prior academic performance, financial status, and religious beliefs. Using a data-mining technique known as CHAID, Chi-Squared Automatic Interaction Detection, Salem found that certain variables such as race, gender, and class are correlated with writing center visits. As a result, Salem suggests that we rethink writing center pedagogy to ensure that it better serves a more diverse population of students. Such a strong conclusion implies the presence of a causal relationship between these factors and writing center usage, yet the statistical technique that the author used to carry out this analysis does not support this claim.

Other writing center studies investigate the effect of attendance on student performance. Agnieszka Bielinska-Kwapisz (2015) analyzes appointment and graded assignment data on 315 undergraduate business students from Montana State University. Using quantile regression, Bielinska-Kwapisz indicates that students in the top 40th percentile who visited the writing center had significantly higher grades on written assignments than those who did not visit the writing center. However, students who were not in the top 40th percentile did not seem to benefit from the writing center. Based on these findings, Bielinska-Kwapisz recommends that writing centers find more ways to aid students in the lower portion of the grade distribution. Although this recommendation seems to follow at first glance, the study does not adjust for confounding variables such as study habits that could themselves explain the difference in performance on writing assignments.

Although Salem and Bielinska-Kwapisz use two well-known statistical methods to arrive at the conclusions mentioned above, both techniques are limited in terms of their ability to identify causal relationships. Methods involving chi-squared tests or regression techniques often ignore unexplained sources of variation, which is a key criterion in establishing causality. As a result, it is possible that potential confounding variables such as sex and race may have impacted Salem and Bielinska-Kwapisz's findings concerning who visits the writing center and the effect of visits on student performance. Hence, in this study we apply a more advanced statistical method rooted in causal inference literature known as propensity score matching to better measure writing center effectiveness.

1.2 Causal Inference Methods–Propensity Score Matching

Randomized controlled trials (RCTs) are considered the "gold standard" approach when it comes to estimating the effect of a treatment on an outcome. As the name suggests, in RCTs subjects are randomly assigned to treatment groups, which ensures that treatment status is not confounded with observed or unobserved baseline characteristics. Therefore, this property of RCTs allows researchers to estimate the effect of a treatment on associated outcomes through direct comparison of the outcomes between treatment groups (Greenland, Pearl, and Robins, 1999).

On the other hand, observational (non-randomized) studies do not enable researchers to directly compare outcomes between treatment groups to estimate treatment effects. This is due to the fact that in observational studies treatment selection is often influenced by subjects' baseline characteristics(Greenland and Morgenstern, 2001). For example, students belonging to one demographic may be more likely to seek out a treatment such as tutoring than other students. Therefore, researchers must account for systematic differences in these baseline characteristics between treatment groups when estimating treatment effects (Rubin, 1997). Typically, researchers use regression adjustment to account for these differences. However, recently there has been increasing interest in a new field of statistics–propensity score methods.

A propensity score is defined as the probability of receiving treatment conditioned on observed covariates. Rosenbaum and Rubin (1983) demonstrate that treated and untreated subjects with the same propensity scores have similar distributions for all covariates. Therefore, this balancing property means that if we control for propensity score when comparing treatment groups, then an observational study effectively becomes a randomized block experiment, in which the blocks are the groups of subjects with the same or extremely similar propensity scores. In essence, this allows us to make causal inferences in observational studies.

Austin (2011b) describes four main propensity score methods: propensity score matching, stratification on the propensity score, inverse probability of treatment weighting using the propensity score, and covariate adjustment using the propensity score. Propensity score matching involves forming matched sets of subjects from different treatment groups who share a similar propensity score. There are several different methods for forming matched pairs, such as optimal, greedy, and caliper-matching. Once a matched sample is created using one of these methods, we can estimate the treatment effect by directly comparing outcomes between subjects in different treatment groups.

Researchers from several different fields have used propensity score matching in their analysis of observational data in order to measure the effect of both binary and multi-level treatments. Yu Ye and Lee Ann Kaskutas (2009) effectively applied propensity score matching to measure the effect of attending Alcoholics Anonymous (AA) meetings, a binary treatment variable, on subject abstinence. Ye and Kaskutas studied 569 patients and found that while AA meetings were more effective for those with a lower propensity to attend the meetings, the program did not make a significant difference for those with a higher propensity to attend meetings. Although the researchers express concern regarding the effect of unobserved factors as a potential limitation of their study, they note that propensity score matching controls for study bias when the data set features a broad collection of covariates. They also cite a key advantage of propensity score matching–the ability to account for all observed covariates without losing statistical power. This suggests that propensity score matching is an effective method for measuring the effect of a binary treatment such as writing center attendance on student performance.

Although propensity score methods are generally confined to binary treatment scenarios, Lu et al (2001) demonstrate how propensity score methods such as matching can be generalized to estimate causal treatment effects in studies involving varying doses of treatment. The researchers illustrate multivariate matching through analysis of data on 521 teens who received varied doses of exposure to an anti-drug media campaign. They use propensity score matching to form 260 high dose—low dose pairs balanced on 22 covariates. Although the researchers found little or no indication of a treatment effect, they demonstrate an effective matching procedure for studies in other fields that involve multi-level treatment variables, such as writing center visits.

Writing is a field that would benefit from a rigorous quantitative study involving this multi-level matching procedure. Salem and Bielinska-Kwapisz both introduce quantitative analysis to the field in their respective studies, yet the statistical methods employed by both researchers prevented either from making definitive conclusions. A more advanced statistical analysis technique that can perhaps address the limitations encountered in these two studies is propensity score matching. Ye and Kaskutas demonstrate that propensity score matching can be used to accurately measure the effect of a binary treatment variable such as attendance, while Lu et al. illustrate how matching can be applied to a multi-level treatment like media-campaign exposure. In this study, we will investigate writing center attendance as a multi-level treatment variable in order to measure the effect of multiple visits on student performance.

The rest of the paper is organized in the following format. First, we provide a thorough description of the process used to carry out this statistical analysis, including everything from data aggregation to propensity score matching. Second, we describe the results of our data analysis, including important data visualizations and tables that estimate the effect of our treatment–writing center visits. Third, we discuss the implications of our results in terms of writing center studies and provide a few limitations of our study that future researchers should consider. Finally, we assess the significance of our study and highlight the key takeaways.

2 Data

The population under observation in this study is students who were enrolled in one of three required, writing-intensive courses at West Point between 2019-2020. The final data set that we use in our propensity score analysis is the aggregate of data from three unique sources. The first data set is 3,351 signature writing event (SWE) grades (percentage) by student and course. The second data set consists of information concerning 22,077 student visits to the MWC, to include the student's name, the date of the visit, and the writing assignment course. The third data set contains background information (sex, race, age, standardized test scores) for each student from institutional admissions records. For students missing SAT or ACT scores, we impute the missing score using a common multivariate imputation approach (Smits, Mellenbergh, and Vorst, 2002). We merge these three data sets into one using a unique student identifier, creating a final aggregated data set composed of 2,662 complete records with the student's SWE grade, an indicator of the student's number of visits to the MWC, and various demographic and academic characteristics concerning the student.

The treatment variable in our study is *MWC Visits*, which represents the number of times that a student visited the writing center for a given SWE. This is a categorical variable with three levels–never, once, multiple times–in which multiple times is defined as more than one visit for a given SWE. It is possible that students visited the MWC for an assignment other than the SWE, yet this is likely rare. The SWEs are heavily weighted (for example, 25% in the introductory composition course) and instructors routinely encourage students to seek assistance from the MWC for these assignments. The response variable is *SWE Grade*, which represents the score that a student received on their signature written event. This is a quantitative variable that ranges from 0 - 100.

Our aggregated data set has the following inclusion criteria. First, we only consider student records of SWEs for the following courses: English Composition (EN101), History of the United States (HI105), International Relations (SS307), and the advanced versions for each of these courses (EN151, HI155, SS357). Other courses did not possess large enough sample sizes to conduct meaningful analysis. Second, we include only student records in which they consulted with either experienced student or professional Writing Fellows for consistency in the quality of appointments. Third, we only consider student records in which they received an event grade of greater than 50%. We found six instances of SWEs that did not meet this criterion, and all of them were completed by students that did not visit the writing center. Assignments scored below this threshold are typically incomplete or late assignments; therefore, they usually indicate a lack of student effort as opposed to poor quality of writing.

3 Methods

Propensity score matching requires a four-step approach that can be carried out using any common statistical programming language such as R (Bryer, 2013). First, we estimate the probability (propensity score) of visiting the writing center for a given writing assignment using logistic regression models. Second, we form matched triplets of subjects from different treatment groups who possess similar propensity scores. In this study, we use caliper-matching in order to specify the greatest potential difference in propensity score between matched subjects. Third, we assess the balance of the measured covariates between the treatment groups, where balance refers to the similarity of the covariate distributions. Finally, we estimate the effect of writing center visits by observing the difference in mean signature writing event scores across the treatment groups amongst matched observations.

The first step, estimating the propensity score, offers the researcher the most discretion. Although there is a lack of consensus in the applied literature with respect to which explanatory variables to include in the estimation models, a recent study demonstrated the merits of including only the true confounders. (Austin, Grootendorst, and Anderson, 2007). Given the data available, we include three explanatory variables in building the logistic regression models: sex, race and standardized test scores. In total, we fit three logistic regression models to assign propensity scores to each observation in our data set. The propensity score indicates the likelihood of a cadet visiting the MWC for a given assignment. The first model estimates the propensity of an individual to attend the MWC multiple times versus once. The second model estimates the propensity of an individual to attend the MWC once versus never. The third model estimates the propensity of an individual to attend the MWC multiple times versus never.

The second step, forming matched triplets of subjects from different treatments, includes a few important modeling decisions as well. There are numerous methods to match observations from different treatment groups. First, we use partial exact matching in order to ensure that all matched observations are within the same course. This accounts for the fact that courses from different academic departments have different established standards and grading rubrics. We then use caliper-matching, wherein the algorithm matches observations from the treatment groups when the distance between their propensity scores is less than a designated caliper. In our study we use a caliper of width equal to 0.25 as recommended by Rosenbaum and Rubin (1985).

The third step, assessing balance, is another important phase of our analysis, as the effectiveness of propensity score methods is dependent on how well balance is achieved. This step involves determining how well our matching algorithm distributed the covariates between treatment groups. There are a few graphical approaches to evaluating balance. For example, covariate balance plots display the distribution of the specified covariate by treatment group using bar charts (categorical variables) and box plots (quantitative variables).

Finally, the fourth step is estimating the effects of writing center visits on signature writing event grades. This requires carrying out t-tests to determine if there is a significant difference in the mean writing assignment grade between the treatment groups. Austin (2011a) recommends using paired t-tests when using propensity-score matched samples for making inferences on the effect of treatment. These tests enable us to measure the effect of visiting multiple times versus once, once versus never, and multiple times versus never.

4 Results

Our analysis accounts for three confounding variables: sex, race, and standardized test scores. **Table 1** included on the following page includes information concerning each of these variables by treatment group. This table indicates that females are overrepresented among the population of students who visited the MWC (once or multiple times). Similarly, this table also illustrates that Asian and African American students are overrepresented in students who visited the MWC. Our imputed results indicate there is not a significant difference between many of these standardized test scores in cadets that went to the MWC and those that did not. However, there does appear to be an association between our treatment variable, *MWC Visits*, and our outcome variable, *SWE Grade*, as those who visited the MWC tended to perform better on SWEs than those who did not.

Of the 2662 records in our aggregated dataset, 95 cadets visited the writing center for an assignment multiple times, 207 cadets visited the writing center for an assignment once, and 2360 cadets never visited the writing center. Our preliminary analysis indicates that those who visited the MWC multiple times received an 83.9% on a given writing assignment and those who visited the MWC once received an 83.0%, whereas those who did not attend received an 81.8%. **Figure 1** included below depicts this graphically. While this seems to suggest that there is a difference between the treatment populations, we did not adjust for any confounding variables. In order to properly adjust for these, we employ propensity score matching.

Table 1: Demographic and Academic Characteristics at Baseline

Characteristic	All (N = 2662)	Multiple Times (N = 95)	Once (N = 207)	Never (N = 2360)	P-Value
Sex–no. (%)					0.000***
Male	2056 (77.2%)	66 (69.5%)	136 (65.7%)	1854 (78.6%)	
Female	606 (22.8%)	29 (30.5%)	71 (34.3%)	506 (21.4%)	
Race–no. (%)					0.037**
Asian	229 (8.6%)	15 (15.8%)	25 (12.1%)	189 (8.0%)	
Black	369 (13.9%)	18 (18.9%)	35 (16.9%)	316 (13.4%)	
Hispanic	243 (9.1%)	4 (4.2%)	14 (6.8%)	225 (9.5%)	
Native American	34 (1.3%)	1 (1.1%)	2 (1.0%)	31 (1.3%)	
Other	40 (1.5%)	2 (2.1%)	3 (1.4%)	35 (1.5%)	
White	1747 (65.6%)	55 (57.9%)	128 (61.8%)	1564 (66.3%)	
SAT–score (200-800)					
Math	652 ± 71.6	661 ± 71.3	652 ± 75.2	651 ± 71.3	0.416
Verbal	617 ± 78.5	612 ± 92.0	614 ± 81.6	617 ± 77.6	0.652
Writing	602 ± 82.9	600 ± 94.1	600 ± 86.1	603 ± 82.2	0.843
ACT–score (1-36)					
Composition	28.6 ± 3.77	28.5 ± 4.17	28.5 ± 3.87	28.6 ± 3.75	0.928
English	28.5 ± 4.89	27.8 ± 5.19	28.5 ± 5.00	28.5 ± 4.87	0.419
Math	28.3 ± 3.66	28.6 ± 3.83	28.3 ± 3.67	28.3 ± 3.65	0.597
Reading	29.3 ± 4.86	29.1 ± 5.97	29.2 ± 4.91	29.3 ± 4.81	0.933
Writing	26.0 ± 4.17	25.7 ± 4.63	26.0 ± 4.22	26.0 ± 4.14	0.811
SWE Grade (0-100)	82.0 ± 8.64	83.9 ± 8.64	83.0 ± 9.06	81.8 ± 8.59	0.014*

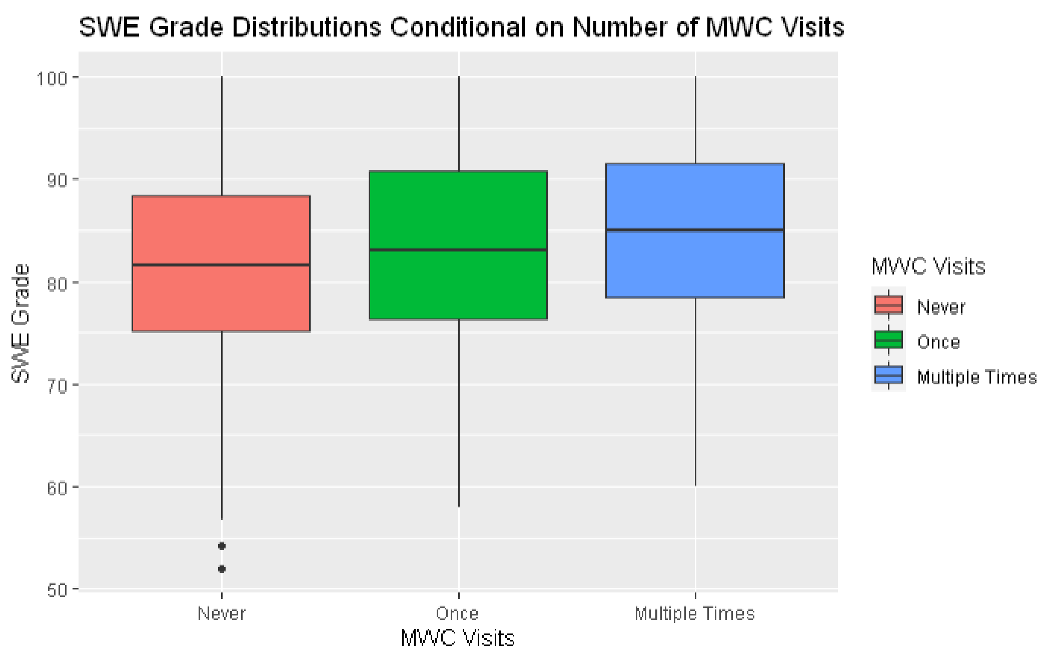


Figure 1: SWE Grade by Number of MWC Visits

4.1 Estimating Propensity Scores

Table 2 included on the following page provides summary information concerning the logistic regression models used to estimate propensity scores for each subject. The first model estimates the propensity of an individual to attend the MWC multiple times versus once. There are three significant predictors included in this model–sex (Male), race (Hispanic), and race (White). Therefore, based on this model, we can gather several useful insights: males are less likely to visit multiple times versus once compared to females, Hispanic students are less likely to visit multiple times versus once, and white students are less likely to visit multiple times versus once. The second model estimates the propensity of an individual to attend the MWC once versus never. There are three significant predictors included in this model–sex (Male), race (Hispanic), and race (White). Therefore, based on this model, we can draw similar conclusions: males are less likely to visit once versus never compared to females, Hispanic students are less likely to visit once versus never, and white students are less likely to visit once versus never. The third model estimates the propensity of an individual to attend the MWC multiple times versus never.

Table 2: Logistic Regression Model Output

	Model 1 (Multiple Visits v. Once)				*Model 2 (Once v. Never)*				*Model 3 (Multiple Times v. Never)*			
Coefficients	**Estimate**	**Std. Error**	**Z Value**	**P Value**	**Estimate**	**Std. Error**	**Z Value**	**P Value**	**Estimate**	**Std. Error**	**Z Value**	**P Value**
Intercept	-1.661	0.869	-1.912	5.589e-02	-3.744	1.273	-2.941	0.003	2.181	1.519	1.435	0.151
Sex (Male)	-0.712	0.165	-4.308	1.650e-05***	-0.654	0.245	-2.666	0.008**	-0.015	0.293	-0.050	0.960
Race (Black)	-0.166	0.307	-0.539	5.896e-01	-0.129	0.410	-0.315	0.753	-0.153	0.517	-0.296	0.768
Race (Hispanic)	-0.774	0.359	-2.159	3.082e-02*	-1.377	0.585	-2.354	0.019*	0.631	0.696	0.907	0.364
Race (Native American)	-0.678	0.770	-0.880	3.788e-01	-0.727	1.068	-0.681	0.496	0.042	1.330	0.032	0.975
Race (Other)	-0.392	0.647	-0.607	5.441e-01	-0.136	0.790	-0.173	0.863	-0.030	1.016	-0.029	0.977
Race (White)	-0.472	0.240	-1.969	4.894e-02*	-0.711	0.315	-2.255	0.024*	0.162	0.408	0.397	0.691
SAT Math	0.002	0.002	0.996	3.191e-01	0.004	0.002	1.528	0.126	-0.001	0.003	-0.422	0.673
SAT Verbal	-0.001	0.002	-0.157	8.755e-01	-0.001	0.002	-0.633	0.527	0.001	0.003	0.311	0.756
SAT Writing	-0.002	0.002	-1.151	2.497e-01	-0.001	0.002	-0.281	0.779	-0.001	0.003	-0.510	0.610
ACT Composition	-0.069	0.105	-0.656	5.118e-01	-0.007	0.153	-0.048	0.962	-0.005	0.191	-0.025	0.980
ACT English	0.021	0.044	0.482	6.298e-01	-0.104	0.065	-1.602	0.109	0.120	0.078	1.532	0.126
ACT Math	0.031	0.050	0.618	5.363e-01	0.047	0.073	0.644	0.520	-0.057	0.092	-0.620	0.535
ACT Reading	0.023	0.042	0.536	5.917e-01	0.048	0.061	0.799	0.424	-0.053	0.076	-0.697	0.486
ACT Writing	0.003	0.039	0.088	9.296e-01	0.030	0.058	0.511	0.610	-0.020	0.075	-0.269	0.788

4.2 Matching

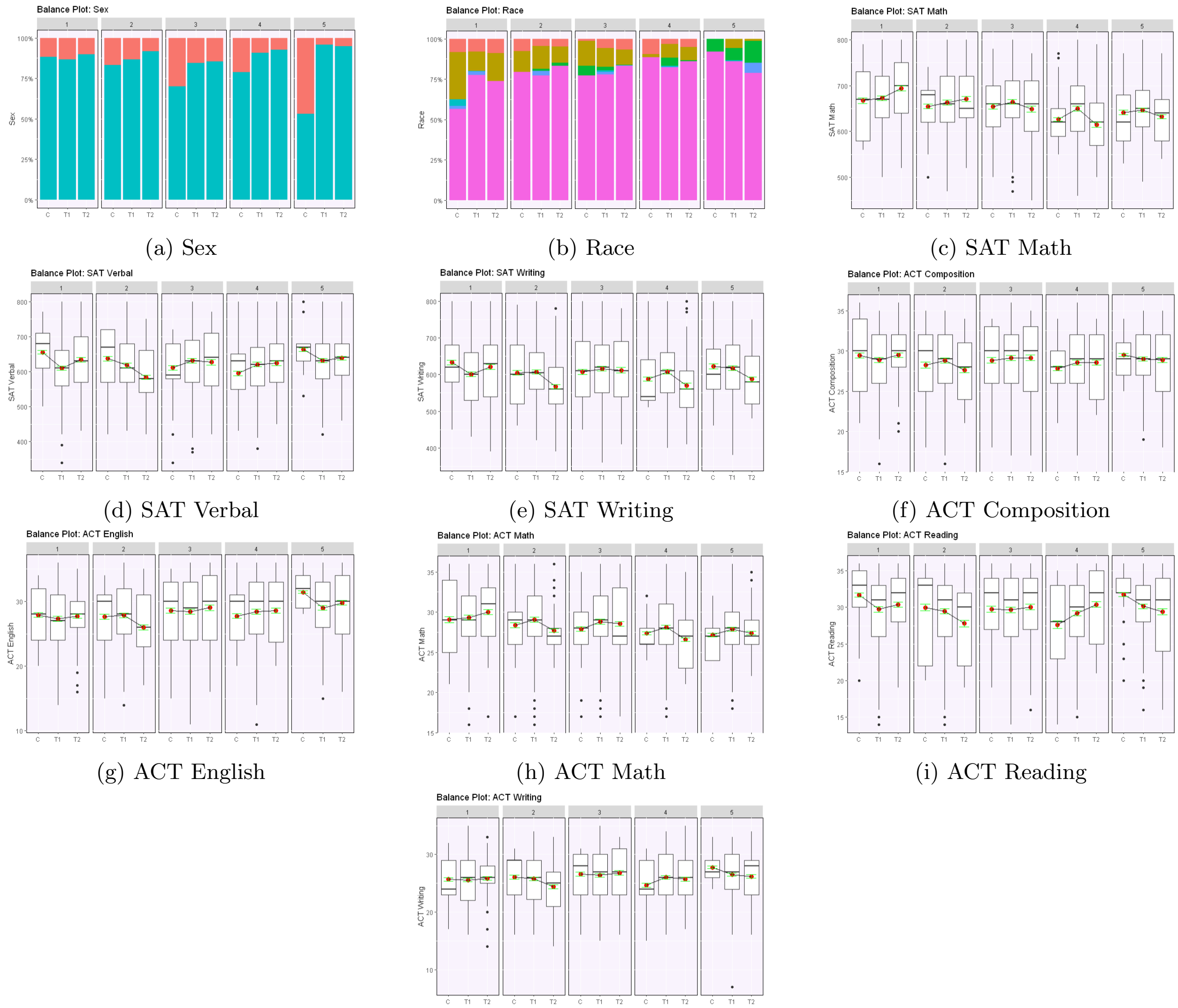
Table 3 shown below provides a summary of the number of unmatched observations by treatment group. Our algorithm matched 1000 (37.6%) of the 2662 observations included in our dataset, leaving 1662 (62.4%) observations not considered in the final comparison of treatment groups. Although the number of unmatched observations may seem large, we must consider that multi-level matching requires forming matched triplets, while binary matching only requires forming matched pairs. It is also worth noting that the majority of these unmatched observations are from the control group (those who never visited the MWC), which is much less concerning than if they had been from the two treatment groups.

Table 3: Unmatched Observations by Treatment

Never	**Once**	**Multiple Times**
1614 (68.4%)	36 (17.4%)	12 (12.6%)

4.3 Assessing Balance

Figure 2 included on the next page consists of balance plots for each covariate included in our analysis. Categorical covariates such as race and sex are represented with bar charts, while quantitative covariates such as standardized test scores are represented with box plots. Our three treatment levels are represented on the x-axis of each plot, with the control group being those who never visited the MWC, treatment one being those who visited once, and treatment two being those who visited multiple times. This figure indicates that the distribution of each covariate is fairly similar across the treatment groups both within and across strata.



(a) Sex (b) Race (c) SAT Math

(d) SAT Verbal (e) SAT Writing (f) ACT Composition

(g) ACT English (h) ACT Math (i) ACT Reading

(j) ACT Writing

Figure 2: Covariate Balance Plots

4.4 Estimating Treatment Effects

Table 4 included on the following page summarizes the results of the three paired t-tests used to determine if there is a significant difference in the mean writing assignment grade between the treatment groups. These t-tests indicate that there is a mean difference of 2.061 percentage points between those who visited the MWC multiple times versus never, 1.619 percentage points between those who visited the MWC multiple times versus once, and 0.443 percentage points between those who visited the MWC once versus never. The mean difference in event grade between treatment groups is depicted visually as well (see **Figure 3** on the next page).

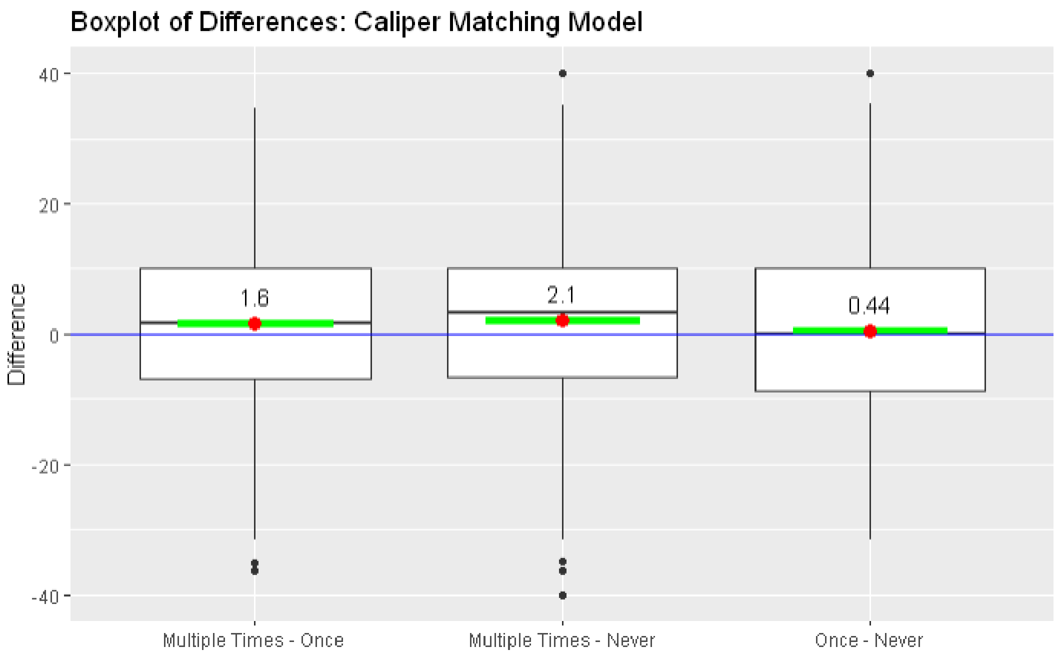


Figure 3: Mean Difference in SWE Grade Between Treatment Groups

Table 4: Estimating Treatment Effects

Treatment	Mean Difference	Confidence Interval	P-Value
Multiple Visits - Never	+2.061	(1.646,2.476)	3.888e-22***
Multiple Visits - Once	+1.619	(1.206,2.031)	1.894e-14***
Once - Never	+0.443	(0.010,0.875)	4.486e-02*

5 Discussion

The study outlined in this report is important given its novelty within the field of writing center studies in terms of both scope and statistical analysis. The final cleaned and merged data set contained two semesters worth of data, resulting in a total of 2,662 observations. This large number of observations provided our analysis with greater statistical power than most other quantitative studies in the field. Not only was our data set relatively large, but it was rich as well. Our final data set consisted of features from three unique sources of data–the MWC, West Point Admissions, and three core curriculum courses. This abundance of features enabled us to adjust for more sources of confounding than most other quantitative writing center studies.

Our study also leveraged a more advanced statistical approach rooted in causal inference literature. In the context of writing center studies, the level of quantitative analysis has been relatively straightforward, with most researchers applying common techniques such as chi-squared tests or regression analysis. On the other hand, propensity score matching is a more advanced method. Though used most to deal with confounding bias in medical research, this study presented an ample opportunity to apply propensity score matching outside the medical field given the baseline differences in the treatment groups. Most researchers use propensity score matching to investigate the effect of a binary treatment variable; however, we chose to treat writing center attendance as a multi-level categorical variable in order to better understand the effect of more visits on performance. This decision to employ multi-level propensity score matching enabled us to identify a significant causal relationship between writing center visits and student performance.

Our results suggest that more visits to the MWC for a core curriculum course resulted in a better grade on the signature writing event within that course. More specifically, the results indicated that visiting the MWC multiple times for a given assignment had a larger effect than only visiting once. This is indicative of a dose-response relationship between visits and performance, as performance increased with each additional number of visits. The existence of such a relationship provides compelling evidence of a causal relationship between MWC visits and SWE grades. Given these findings, the directors of the MWC should consider altering their advertising approach in order to attract clients to schedule multiple appointments for their written assignments. Considering the standard liberal arts curriculum studied at the Academy, these findings are relatively generalizable to writing centers at other institutions as well, especially liberal arts schools. Therefore, these findings should help inform decision making within the MWC and other undergraduate writing centers across the nation.

There are a few limitations to our analysis. First, there are likely confounders that we did not adjust for, and thus the effect that we measured may not be entirely attributable to writing center visits. In building the logistic regression models to estimate subjects' propensity scores, we included three predictor variables–sex, race, and standardized test scores. Although it is impossible to ensure that there is no unmeasured confounding in these models, there are likely other predictors that would improve its accuracy. However, the data set used in our study did not include other potential predictors such as the subject's academic major.

Second, our analysis was directed at measuring the effect of visits on assignment grades opposed to writer quality. The general consensus in the field of writing center studies is that writing center visits are not meant to produce better papers–they are meant to produce better writers. Although writing center staff members are certainly pleased when clients perform better on their written assignments, their overarching mission is to help individuals become more effective written communicators in the long term. Therefore, this study offers only limited insight in terms of writing center performance towards accomplishing this goal.

Given the limitations highlighted throughout this section, there are many opportunities for future research. One particularly useful study would involve measuring the cumulative effect of regularly visiting the writing center over time. Such a study would perhaps be a more appropriate indicator of writing center performance towards creating better writers than a related outcome such as SWE grade. Furthermore, future studies should also seek to analyze more confounding variables than those included in this one. For instance, one variable that we were not able to analyze in our study due to a lack of data that would potentially be informative is the subject's academic major. Including this variable in the study may explain more variation in the observed outcome, as it is reasonable to assume that humanities majors may be both more inclined to visit the writing center and more inherently talented at writing than a technical major.

Overall, despite the two limitations highlighted throughout this section, our study provides an effective statistical analysis of the effect of visiting a writing center on student grades for signature writing events in core curriculum courses. Other existing writing center studies have attempted to do the same, yet with less scientific rigor. Our study introduces quantitative rigor to the field through the use of multi-level propensity score matching–an advanced statistical method that is new to writing center studies. Our results indicate that more visits for a core curriculum course resulted in a better grade on the signature writing event within that course. More specifically, we found a dose-response relationship between writing center visits and student performance on signature writing events. Ultimately, these results will allow the MWC and other undergraduate writing centers to make programmatic decisions based on data-driven evidence rather than intuition.

6 Appendix A: Assessing Balance

In the body of this report we include balance plots in order to assess the balance of the measured covariates between the treatment groups. The following table provides another method to assess balance.

Table 5: Balance Estimates

Covariate	Friedman Test	P Value	rmANOVA	P Value
Sex	573.947	2.339e-125***	NA	NA
Race	7.076	2.907e-02*	NA	NA
SAT Math	36.255	1.341e-08***	22.018	2.945e-10***
SAT Verbal	22.094	1.5934e-05***	18.920	6.401e-09***
SAT Writing	96.796	9.572e-22***	57.636	1.523e-25***
ACT Composition	0.517	7.720e-01	1.068	3.438e-01
ACT English	7.146	2.807e-02*	10.015	4.541e-05***
ACT Math	60.930	5.8778e-14***	34.278	1.547e-15***
ACT Reading	56.193	6.277e-13***	13.539	1.355e-06***
ACT Writing	4.287	1.173e-01	7.492	5.623e-04***

References

[Austin, 2011a] Austin, P. C. (2011a). Comparing paired vs non-paired statistical methods of analyses when making inferences about absolute risk reductions in propensity-score matched samples. *Statistics in medicine*, 30(11):1292–1301.

[Austin, 2011b] Austin, P. C. (2011b). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, 46(3):399–424.

[Austin et al., 2007] Austin, P. C., Grootendorst, P., and Anderson, G. M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a monte carlo study. *Statistics in medicine*, 26(4):734–753.

[Bielinska-Kwapisz, 2015] Bielinska-Kwapisz, A. (2015). Impact of writing proficiency and writing center participation on academic performance. *International Journal of Educational Management*.

[Bryer, 2013] Bryer, J. M. (2013). Trimatch: An r package for propensity score matching of non-binary treatments. In *The R user conference, useR*, pages 10–12. Citeseer.

[Greenland and Morgenstern, 2001] Greenland, S. and Morgenstern, H. (2001). Confounding in health research. *Annual review of public health*, 22(1):189–212.

[Greenland et al., 1999] Greenland, S., Pearl, J., and Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, pages 37–48.

[Lu et al., 2001] Lu, B., Zanutto, E., Hornik, R., and Rosenbaum, P. R. (2001). Matching with doses in an observational study of a media campaign against drug abuse. *Journal of the American Statistical Association*, 96(456):1245–1253.

[Rosenbaum and Rubin, 1983] 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.

[Rosenbaum and Rubin, 1985] Rosenbaum, P. R. and Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1):33–38.

[Rubin, 1997] Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of internal medicine*, 127(8_Part_2):757–763.

[Salem, 2016] Salem, L. (2016). Decisions... decisions: Who chooses to use the writing center? *The Writing Center Journal*, pages 147–171.

[Smits et al., 2002] Smits, N., Mellenbergh, G. J., and Vorst, H. C. (2002). Alternative missing data techniques to grade point average: Imputing unavailable grades. *Journal of Educational Measurement*, 39(3):187–206.

[Ye and Kaskutas, 2009] Ye, Y. and Kaskutas, L. A. (2009). Using propensity scores to adjust for selection bias when assessing the effectiveness of alcoholics anonymous in observational studies. *Drug and Alcohol Dependence*, 104(1-2):56–64.