Praxis: A Writing Center Journal • Vol. 19, No. 3 (2022)

Measuring the Effect of Writing Center Visits on Student Performance

Joseph Zuccarelli
Harvard University
joseph_zuccarelli@g.harvard.edu

Nicholas Cunningham
United States Military Academy
nicholas.cunningham@wespoint.edu

Colleen G. Eils
United States Military Academy
colleen.eils@wespoint.edu

Andrew Lee
United States Military Academy
andrew.lee@westpoint.edu

Kevin Cummiskey
United States Military Acdemy
kevin.cummiskey@westpoint.edu

Abstract

Given the recent rise of data science, a growing number of scholars are publishing quantitative studies on the impacts of writing centers. Typically, these studies aim to answer assessment-style questions such as "Who visits the writing center?" and "Are writing center visits effective in terms of increasing student performance?" The majority of these studies feature the application of common statistical approaches such as correlation and regression analysis, which provide useful but limited results. In this article, we apply a more complex statistical method known as propensity score matching in order to identify factors associated with writing center attendance as well as the effect of visits on student written performance. In total, we analyzed two semesters of visits to the Mounger Writing Center (MWC) at the United States Military Academy at West Point and over 2,500 student records of signature writing assignments. We found statistically significant evidence that race and gender are associated with attendance, specifically that women and historically underrepresented students are more likely to visit the writing center for signature writing assignments. We identified the presence of a causal relationship between MWC visits and student grades on signature writing assignments, as those who visited the MWC multiple times received grades approximately 2% greater on average compared to those who did not visit. Ultimately, this article provides administrators with a more robust quantitative framework to assess the efficacy of their writing center, thus enabling for more informed programmatic decisions.

INTRODUCTION

Undergraduate consultants at the Mounger Writing Center (MWC) at the United States Military Academy (USMA), better known as West Point, are relatively distinct in two ways: we staff the MWC as part of academic coursework, rather than for pay, and as a group we major in programs from every department on campus. Roughly a third humanities, social sciences, and STEM majors, respectively, we bring uniquely multidisciplinary perspectives to the seminar courses we take in the pedagogy and practices of college writing and to our writing center consultations. The interdisciplinary connections that start as part of our writing fellows courses, which are grounded in foundational and contemporary scholarship in writing studies and writing center studies, often extend beyond the two-course sequence and into our respective disciplinary research interests. This was the case for me, the first author.

As a mathematical sciences major and a cadet writing fellow, as our undergraduate consultants are called, I quickly became interested in quantitative writing center studies, specifically those aimed at assessing the efficacy of writing centers. Although qualitative methods of study are more prevalent, a few key scholars have used quantitative methodologies in their assessments of writing centers. Recently, Lori Salem made waves within the field with the article “Decisions...Decisions: Who Chooses to Use the Writing Center,” in which she studied the differences between users and non-users of the writing center at Temple University through correlation analysis. Salem was not the first scholar to carry out a quantitative study, however, as others such as Neal Lerner and Stephen Newmann did as well a few decades before. Both Lerner and Newmann sought to evaluate the effect of visiting the writing center by comparing grades of students who used the writing center versus those who did not, positioning students at similar starting points by using SAT Verbal scores.

The three studies mentioned above all involved quantitative approaches aimed at answering two key questions within the field: (1) What factors affect writing center attendance? (2) What is the effect of writing center visits on students' performance on written assignments? The answers to these two questions enable writing center administrators to assess the effectiveness of their services, as they speak to a center's ability to attract clients as well as the quality of visits. Although the traditional purpose of a writing center is clear, "to produce better writers, not better writing," scholars have yet to agree upon the best way to assess centers' ability to achieve this objective (North 438). Regardless, writing center administrators must assess something—student satisfaction, students' self-reported learning outcomes, writers' confidence levels—to highlight their successes, justify program budgets, and drive pedagogical innovations (Lape 1). In this article, we offer a more complex statistical framework known as propensity score matching that can provide writing center administrators with more accurate causal inferences using data-driven answers to these typical assessment style questions concerning writing center efficacy.

The rest of the article is organized in the following format. First, we highlight a few noteworthy quantitative studies aimed at assessing the effectiveness of writing centers. Second, we describe the process used to carry out our statistical analysis, from data aggregation to propensity score matching. Third, we lay out the results of our data analysis, including important data visualizations and tables that indicate the major factors concerning writing center attendance and the effect of attendance on students' written performance. Fourth, we discuss the implications of our results in terms of writing center studies and identify a few limitations of our study that future researchers should consider. Finally, we assess the significance of our study as a framework or statistical approach that other writing centers can follow.

LITERATURE REVIEW: QUANTITATIVE WRITING CENTER STUDIES

Conventionally, authors of writing center studies use qualitative methodologies such as focus groups, case studies, and surveys, in their assessments of writing centers (Gofine 41). Increasingly, however, scholars and administrators have sought out quantitative methods to provide a different perspective on writing center operations. Thus far, these quantitative methods have largely ranged anywhere from basic descriptive statistics all the way to correlation and regression analysis (Newman, Lerner, Salem, Bielinska-Kwapisz).

Descriptive statistics are simple quantifications commonly used by writing center administrators to report matters such as how many students visited their writing center, each student's major, and the number of tutoring hours completed by each consultant. Although these numbers are often a necessary component of writing centers' annual reports, administrators cannot rely on them to assess the efficacy of their services. Such basic statistics may be used to gauge whether or not a center is attracting clients; however, they in no way indicate the quality of services that clients receive during visits (Bell 9). Therefore, in order to draw useful insights from data, scholars have turned to more refined quantitative analysis methods.

Two of the first scholars to carry out such analysis within the field were Newmann and Lerner. In

“Demonstrating Effectiveness” and “Counting Beans and Making Beans Count,” Newman and Lerner, respectively, sought to determine if students who use the writing center achieved higher first-year composition (FYC) grades than those who did not. To do so, the two scholars used a similar methodology—compare the grades of those students who used the writing center with those who did not, while adjusting for students’ SAT verbal scores. Both Newmann and Lerner reported similar results. Newmann found that significantly fewer students who visited the writing center received unsuccessful grades than those who did not visit (8). Likewise, Lerner claimed that those with lower SAT verbal scores benefited most from visits, as this group performed about five percentage points better on average than those in the same SAT range who did not visit the writing center (“Counting Beans and Making Beans Count” 3).

Although these two researchers used similar methods and found similar results, Lerner later acknowledged that both studies were not statistically sound. According to Lerner, both studies were grounded in three invalid assumptions. First, both assumed that students with lower SAT verbal scores were at a disadvantage in FYC courses. Lerner found at his institution the relationship between students’ SAT verbal scores and FYC grades was extremely weak, even weaker than the relationship between SAT math scores and FYC grades ("Writing Center Assessment" 62). Second, both assumed that a student’s final FYC grade was a proper indication of his or her writing ability. This assumption is also troublesome, as we cannot be certain that a student’s final course grade is a fair assessment of the goals that a writing center holds for its visitors. As mentioned previously, the purpose of a writing center is to improve students’ writing ability, and FYC grades also measure student's diligence in class participation and ability to complete assignments on time. Third, both assumed that grading was consistent across FYC sections. Lerner found at his institution that instructors were not using the same grading criteria, thus invalidating this assumption ("Writing Center Assessment " 63). Given these flaws, Lerner concluded that both his and Newmann's findings were unreliable.

Recently, several scholars carried out studies with similar objectives using more traditional statistical approaches. For instance, in "Impact of Writing Proficiency and Writing Center Participation on Academic Performance," Agnieszka Bielinska-Kwapisz analyzed data for appointments and graded assignments from 315 undergraduate business students at Montana State University to investigate the effect of writing center attendance on student performance (385). Using regression analysis, Bielinska-Kwapisz indicated that students in the top 40th percentile of the grade distribution who visited the writing center had significantly higher grades on written assignments than those who did not visit the writing center. However, students below this threshold did not seem to benefit from the writing center (391). Based on these findings, Bielinska-Kwapisz recommended that writing centers find more ways to aid students in the lower portion of the grade distribution (392). Although this recommendation seems to follow at first glance, the study does not adjust for all possible confounding variables such as study habits that could themselves explain the difference in performance on writing assignments. For example, those students inclined to visit the writing center may have been disproportionately inclined to start their assignments well in advance.

Similarly, Lori Salem studied factors associated with writing center attendance using data from 4,202 students at Temple University. Salem analyzed several potential factors associated with writing center attendance such as prior academic performance, beliefs and preferences, and financial status. Using a data-mining technique known as CHAID, or Chi-Squared Automatic Interaction Detection, Salem found that certain variables such as race, gender, and class were correlated with writing center visits. Specifically, students most likely to visit the writing center were those "historically excluded from full access to higher education: women, students of color, English language learners, and students with less inherited merit" (160). As a result, Salem suggested that we fundamentally rethink writing center pedagogy to ensure that it better serves a more diverse population of students. Although this suggestion seems to follow from Salem's analysis, this study is limited in that Salem only identifies associations between variables.

Although these authors used well-known statistical methods to arrive at their conclusions, both studies are limited in terms of their ability to identify causal relationships. Methods involving correlation or regression techniques often ignore unexplained sources of variation, which is a key criterion in establishing causal relationships. As a result, it is possible that other variables such as gender, race, or prior academic performance may impact findings concerning the effect of visits on student performance. Hence, in this study we apply a method rooted in causal inference literature known as propensity score matching to better measure writing center effectiveness.

DATA

The population under observation in this study is students who were enrolled in Freshman Composition, US History, or International Relations—three of the more writing-intensive core courses required of all students at USMA—between 2019-2020. Importantly, all three courses are linked to the West Point Writing Program, which requires signature writing events (SWEs) as part of its Writing-Across-the-Curriculum approach to developing undergraduate writers. This means that each of these courses has a standard, course-wide writing event, which, in accordance with the principles of "signature work" as outlined by the American Association of American Colleges and Universities, focuses students' attention on a particular assignment.[1] The final data set used in our statistical analysis is the aggregate of data from three unique sources. The first data set is 3,351 SWE grades by student and course. The second data set consists of information concerning 1,389 student visits to the MWC, to include the student's name, the date of the visit, and the course of the assignment for which they are visiting the writing center. The third data set contains background information available (gender, race, standardized test scores) for 6,817 students from institutional admissions records.[2] 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.[3]

The variable whose effect we are interested in measuring, or the "treatment variable," is *MWC Visits*, which represents the number of times that a student visited the writing center for a given SWE. *MWC Visits* is a categorical variable with three treatment groups—never, once, multiple times—in which multiple times is defined as more than one visit for a given SWE. Students at any stage in the writing process are encouraged to visit the MWC. Prior to any visit, students are required to fill out an appointment form that indicates, among other elements, their preferred focus for the visit. Visits are typically 45 minutes in length, during which writing fellows engage clients in productive conversations about their ideas and how to express them more effectively.

The variable that we are observing in order to determine the effect of our treatment, or the "outcome variable," is *SWE Grade*, which represents the score that a student received on their signature writing event. *SWE Grade* is a quantitative variable that ranges from 0 - 100. The SWE represents a significant writing assignment in the course, as determined by the program or course director. They usually occur in the latter stages of a course, providing time for students to develop as thinkers and writers within the given discipline. SWEs are prepared writing assignments, meaning students are provided time to engage in a full writing process. Each of the assignments included in this study is formally scaffolded and affords students weeks to plan, draft, and revise the essay. All instructors within each course are required to use the same criteria and grading scales to assess students' SWEs. While they may not achieve perfect interrater reliability, each of the core courses represented within our dataset run regular calibration exercises for faculty.

Our aggregated data set has the following inclusion criteria. First, we only consider student records of SWEs for the following courses: First-Year Composition, History of the United States, International Relations, and the advanced versions for each of these courses. Other courses did not possess large enough sample sizes to conduct meaningful analysis. Second, we only consider records of students who received a SWE grade of greater than 50%. We found six instances of SWEs that did not meet this criterion, and all of them came from students who did not visit the writing center. SWEs scored below this threshold are typically incomplete or late assignments; therefore, these grades may not indicate poor quality of writing.

METHODS

Randomized controlled trials (RCTs) are considered the “gold standard" approach when it comes to estimating the effect of a treatment, such as a writing center visit, on an outcome, such as a student’s SWE grade. In RCTs, as the name suggests, subjects are randomly assigned to treatment groups, which simply represent the degree of treatment that they must receive. This property of RCTs allows researchers to estimate treatment effects through direct comparison of the outcomes between treatment groups (Greenland, Pearl, and Robins, 1999). Although arguably every writing center administrator would be interested in estimating the effect of visits to their center on students’ written performance, it would be unfair to mandatorily assign students at random to a certain number of visits. Furthermore, the pedagogy of many writing centers requires that students come of their own accord, which is fundamentally at odds with making random assignments as part of an experiment. Therefore, administrators are limited to another quantitative approach known as an observational study.

Observational studies are those in which the researcher observes individuals without any form of intervention, oftentimes due to ethical concerns or logistical constraints. Unlike RCTs, these studies do not enable researchers to directly compare an outcome between treatment groups to estimate the effect of a treatment. This is due to the fact that in observational studies treatment selection is often influenced by subjects' baseline characteristics, meaning that students of a specific gender, race or standardized test score range may be more likely to seek out a treatment such as visiting the writing center than other students (Greenland and Morgenstern 193). In statistics, this is known as confounding, which occurs when the set of variables that predispose subjects to receive a certain degree of treatment are also related to the outcome. Researchers must properly account for potential sources of confounding in order to accurately estimate treatment effects (Greenland and Morgenstern 199). One method to account for these differences is propensity score matching.

Propensity score matching (PSM) is a statistical technique used in observational studies to reduce bias caused by confounding in order to more accurately estimate the effect of a treatment on an outcome. For example, in our study we use PSM in order to estimate the effect of visiting the MWC on student SWE grades. However, we cannot estimate the effect of visits by directly comparing the SWE grades of those who did and did not visit the MWC, as this estimate would be biased by variables that are associated with the number of times that a student visits and their SWE grade (e.g.: gender, race, ACT/SAT scores). These variables are typically referred to as confounding variables, as they influence both the treatment and the outcome. For example, students who were high performers on standardized tests might visit the writing center more often and also earn a higher score on their SWE than the average student. PSM attempts to control for confounding variables by making the groups receiving treatment and those not receiving treatment comparable with respect to these variables.

PSM requires a three-step approach that can be carried out using any common statistical programming language such as R, which is open source and freely available (Bryer). First, we estimate the probability, i.e., propensity score, that a student visits the MWC for a given SWE based on their gender, race, and standardized test scores. There are a few different methods to carry out this estimation—we use a common statistical technique known as the logistic regression model.[4] Second, we match students from different treatment groups (*MWC Visits*—never, once, multiple times) who possess similar propensity scores.[5] Note that it is important to evaluate these matchings to ensure that the confounding variables included in our model—gender, race, standardized test scores—are evenly distributed amongst the matched groups. Finally, we assess the effect of writing center visits by observing the difference in mean SWE grades across the treatment groups amongst matched observations. This requires carrying out t-tests to determine if there is a significant difference in the mean SWE grade between the treatment groups.[6] These tests enable us to measure the effect of visiting multiple times versus once, once versus never, and multiple times versus never.

RESULTS

Prior to our PSM approach, we first carried out some exploratory analysis of our dataset. Of the 2,662 records in our dataset, 95 students visited the MWC for an assignment multiple times, 207 students visited the MWC for an assignment once, and 2,360 students never visited the MWC. Table 1 (see Appendix A) provides a breakdown of students' academic and demographic characteristics by number of MWC visits. This table indicates that, in terms of gender, women are overrepresented among the population of students who visited the MWC once or more (16.5%), and in terms of race, Asian American (17.5%) and African American (14.4%) students are disproportionately likely to visit the MWC. This table also indicates that there is not a significant difference in standardized test scores between students that went to the MWC and those that did not.

Figure 1 (see Appendix B) provides a visual indication of students' SWE grades by the number of times they visited the MWC. Note that those who visited the MWC multiple times received an average of 83.9% on a given SWE and those who visited the MWC once received an 83.0%, whereas those who did not attend received an 81.8%. Although this direct comparison across our treatment groups suggests that more visits to the MWC is associated with greater performance on SWEs, we did not adjust for any of the confounding variables mentioned previously. Therefore, while it appears that visits to the MWC are improving students' SWE performance, we cannot determine if this enhanced SWE performance is entirely attributable to MWC visits or if there are other variables such as students' gender, race and/or standardized test scores that are confounding our findings. In order to properly adjust for these confounding variables, we employ PSM.

Recall that the first step in our PSM approach is to use logistic regression models to estimate the propensity of each student to visit the MWC given their demographic and academic characteristics (gender, race, standardized test scores). Table 2 (see Appendix A) contains summary information concerning each of these models. These models provide several useful insights concerning the factors affecting attendance at the MWC. Specifically, we find a statistically significant association between gender and MWC visits, as men are less likely to visit the MWC multiple times versus once compared to women. Similarly, we also find a statistically significant association between race and MWC visits, as Hispanic students are less likely to visit multiple times versus once, and white students are less likely to visit multiple times versus once. These same conclusions all hold when comparing those who visit once versus never as well: men are less likely to visit once versus never compared to women, Hispanic students are less likely to visit once versus never, and white students are less likely to visit once versus never.

Moving along in our three-step PSM approach, the second step is to match students with similar propensity scores across the three treatment groups: never, once, and multiple times. In order to ensure quality matchings, our algorithm only matched observations from the same course for which the difference in propensity scores fell below a given threshold (see Endnote 5). Using this threshold, our algorithm matched 1,000 (37.6%) of the 2,662 observations included in our dataset, leaving 1,662 (62.4%) observations not considered in the final comparison of treatment groups. Table 3 (see Appendix A) provides a full breakdown of unmatched observations by treatment group. Although the number of unmatched observations may seem large, we must consider that the majority (68.4%) of these unmatched observations came from those who never visited the MWC, which is much less concerning than if they had been from the two treatment groups (those who visited the MWC one or multiple times). Figure 2 (see Appendix B) consists of balance plots that provide a visual indication of the quality of our matching procedure.

The third and final step in our PSM approach is to use t-tests to estimate the effect of visiting the MWC on students' SWE grades. Table 4 (see Appendix A) summarizes the results of the three paired t-tests used to determine if there is a significant difference in the mean SWE grade between 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. Figure 3 (see Appendix B) graphically depicts the mean differences between each treatment group. Note that all three mean differences were statistically significant.

DISCUSSION

The study outlined in this report is significant given its novelty within the field of writing center studies in terms of both scope of data 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 rigor than most other quantitative studies in the field. Not only was our data set relatively large, but it was also rich, as it consisted of features from three unique sources of data—the MWC, USMA Admissions, and three core curriculum courses. This aspect of our data set enabled us to adjust for more sources of confounding than most other quantitative writing center studies, leading to a more accurate estimate of the effect of writing center visits on students' SWE grades.

Most importantly, our study utilized a more complex analytical approach in propensity score matching. In the context of writing center studies, the level of quantitative analysis has been relatively straightforward, with most researchers applying standard techniques such as chi-squared tests or regression analysis. Although these techniques can certainly be useful, they are limited in that they often can only quantify how changes in a treatment are associated with changes in an outcome. On the other hand, propensity score matching is a more powerful method, as it better controls for confounding variables than traditional techniques in order to produce a more accurate estimate of the treatment effect on an outcome. This methodology enabled us to identify a significant causal relationship between writing center visits and student performance.

Our propensity score matching approach, when tailored to an institution's specific contexts, is advantageous in that it provides evidence-backed answers to two key questions for writing center administrators: (1) Who visits the writing center? (2) Are writing center visits effective in terms of student performance on signature writing assignments? Our analysis provides several valuable insights concerning the demographic and academic characteristics of those who frequent West Point's MWC. In terms of gender, our results suggest that women are more likely than men to visit the MWC once or multiple times. Similarly, in terms of race, our results suggest that Asian and African American students are more likely to visit than students of other racial backgrounds. Given these findings, MWC administrators can now alter or reinforce their advertising and pedagogy appropriately.

Our analysis also provides MWC administrators with informative results regarding the effectiveness of visits, as our findings suggest that more visits to the MWC for a core curriculum course resulted in a better grade on the SWE within that course. Specifically, we found that visiting the MWC multiple times for a given SWE had a larger effect than only visiting once. This is further indicative of a significant causal relationship between MWC visits and SWE performance, as performance increased with more visits. Knowledge of this relationship can be used to inform major administrative decisions within the MWC.

Although these findings are both accurate and informative, there are a few limitations to our analysis. First, there are likely confounding variables for which we did not adjust, and thus the effect that we measured on student SWE grades may not be entirely attributable to writing center visits. While it is impossible to ensure that there is no unmeasured confounding in these models, there still might be other variables that would improve the resulting estimates. For instance, future studies should consider including students’ instructors and academic majors in their statistical models, as this may explain more variation in the observed outcome, SWE grade.

Second, our analysis was directed at measuring the effect of visits on assignment grades as opposed to writer development. The general consensus in the field of writing center studies is that the objective of visits is to ensure that writers, not necessarily their papers, are improving. 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, future studies should seek to measure the cumulative effect of regularly visiting the writing center over time. This would perhaps provide a better indication of writing center performance towards creating better writers than a related outcome such as SWE grade.

Overall, our study provides an effective statistical analysis on the population who attends the writing center and the effectiveness of visits in terms of student performance on signature writing assignments. 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 propensity score matching—a more complex statistical method that had yet to be applied to writing center studies. Although the results of our study are specific to the MWC at West Point, the methodology and framework used to produce these results can be replicated to carry out a similar study involving data from any writing center, accounting for specific institutional contexts. Such a study would provide other writing center administrative teams with useful insights concerning attendance as well as visit effectiveness. Ultimately, studies such as this one will enable undergraduate writing centers to make more informed programmatic decisions that are based on data-driven evidence.

NOTES

1. The Association of American Colleges and Universities defines “signature work” as an individual project related to a significant issue, problem, or question that students’ define for themselves—immersing themselves in exploration, applying what they learn to real-world situations, and preparing to explain the significance of their work to others (“Integrative Learning and Signature Work”).

2. This data set contains both ACT and SAT scores for each student. USMA does not require its applicants to submit scores for both tests; therefore, we infer any missing scores using a common multivariate imputation approach (Smits et al.).

3. This study was approved by and complied with the regulations of the Institutional Review Board (IRB).

4. In total, we fit three logistic regression models to assign propensity scores to each observation in our data set. These models estimate the probability of attending the writing center multiple times versus once, once versus never, and multiple times versus never, given a student’s gender, race, and standardized test scores (SAT and ACT).

5. There are several different matching techniques. In this study we 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. Our algorithm uses a caliper of width equal to 0.25, as recommended by Rosenbaum and Rubin (37). Our algorithm also ensures that matched observations are within the same course, as this accounts for the fact that courses from

different academic departments have different established standards and grading scales.

6. Austin recommends using paired t-tests when using propensity-score matched samples for making inferences on the effect of a treatment (1298).

WORKS CITED

Austin, Peter C. "Comparing Paired vs Non-paired Statistical Methods of Analyses when Making Inferences About Absolute Risk Reductions in Propensity-Score Matched Samples." *Statistics in Medicine*, vol. 30, no.11, 2011, pp. 1292-1301. https://doi.org/10.1002/sim.4200.

Bell, James H. "When Hard Questions are Asked: Evaluating Writing Centers." *The Writing Center Journal* 21.1 (2000): 7-28. https://www.jstor.org/stable/43442109.

Bielinska-Kwapisz, Agnieszka. "Impact of Writing Proficiency and Writing Center Participation on Academic Performance." *International Journal of Educational Management*, 2015. https://doi.org/10.1108/IJEM-05-2014-0067.

Bryer, Jason M. “TriMatch: An R Package for Propensity Score Matching of Non-Binary Treatments.” *The R User Conference, Use R*, 2013, pp. 10-12. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.398.4972&rep=rep1&type=pdf

Gofine, Miriam. "How are we Doing? A Review of Assessments within Writing Centers." *The Writing Center Journal* 32.1 (2012): 39-49. https://www.jstor.org/stable/43442380.

Greenland, Sander, Judea Pearl, and James M. Robins. "Causal Diagrams for Epidemiologic Research." *Epidemiology*, 1999, pp. 37-48. https://www.jstor.org/stable/3702180.

Greenland, Sander, and Hal Morgenstern. "Confounding in Health Research." *Annual Review of Public Health*, vol. 22, no. 1, 2001, pp. 189-212. https://www.annualreviews.org/doi/full/10.1146/annurev.publhealth.22.1.189.

"Integrative Learning and Signature Work." *American Association of Colleges and Universities*, 2022. https://www.aacu.org/office-of-global-citizenship-for-campus-community-and-careers/integrative-learning.

Lape, Noreen. "The Worth of the Writing Center: Numbers, Value, Culture, and the Rhetoric of Budget Proposals." *Praxis: A Writing Center Journal*, vol. 10, no. 1, 2012, pp. 1-4. http://hdl.handle.net/2152/62143.

Lerner, Neal. "Counting Beans and Making Beans Count." *Writing Lab Newsletter*, vol. 22, no. 1, 1997, pp. 1-3. https://wlnjournal.org/archives/v22/22-1.pdf.

Lerner, Neal. "Writing Center Assessment." *Hold: Critical Perspectives on Writing Center Scholarship,* edited by M. Pemberton, & J. Kinkead (2003): 58-73. https://www.jstor.org/stable/j.ctt46nxnq.

Newmann, Stephen. “Demonstrating Effectiveness.” *Writing Lab Newsletter*, vol. 23, no. 8, 1999, pp. 8-9. https://www.wlnjournal.org/archives/v23/23.8.pdf.

North, Stephen M. “The Idea of a Writing Center.” *College English*, vol. 46, no. 5, 1984, pp. 443-446. https://www.jstor.org/stable/377047.

Rosenbaum, Paul R., and Donald B. Rubin. "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score." *The American Statistician*, vol. 39, no. 1, 1985, pp. 33-38. https://doi.org/10.2307/2683903.

Salem, Lori. “Decisions....Decisions: Who Chooses to use the Writing Center?” *The Writing Center Journal*, 2016, pp. 147-171. https://www.jstor.org/stable/43824060.

Smits, Niels, Gideon J. Mellenbergh, and Harrie CM Vorst. "Alternative Missing Data Techniques to Grade Point Average: Imputing Unavailable Grades." *Journal of Educational Measurement*, vol. 39, no. 3, 2002, pp. 187-206. https://doi.org/10.1111/j.1745-3984.2002.tb01173.x.

APPENDIX A

Table 1. Demographic and Academic Characteristics at Baseline. °

Characteristic	All (N = 2662)	Multiple Times (N = 95)	Once (N = 207)	Never (N = 2360)
Gender—no. (%) ***				
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. (%) ***				
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 (65.5%)	55 (57.9%)	128 (61.8%)	1564 (66.3%)
SAT—score (200 - 800) Ɨ				
Math	652 ± 71.6	661 ± 71.3	652 ± 75.2	651 ± 71.3
Verbal	617 ± 78.5	612 ± 92.0	614 ± 81.6	617 ± 77.6
Writing	602 ± 82.9	600 ± 94.1	600 ± 86.1	603 ± 82.2
ACT—score (1 – 36) Ɨ				
Composition	28.6 ± 3.77	28.5 ± 4.17	28.5 ± 3.87	28.6 ± 3.75
English	28.5 ± 4.89	27.8 ± 5.19	28.5 ± 5.00	28.5 ± 4.87
Math	28.3 ± 3.66	28.6 ± 3.83	28.3 ± 3.67	28.3 ± 3.65
Reading	29.3 ± 4.86	29.1 ± 5.97	29.2 ± 4.91	29.3 ± 4.81
Writing	26.0 ± 4.17	25.7 ± 4.63	26.0 ± 4.22	26.0 ± 4.14
SWE Grade—score (0 -100) *	***82.0 ± 8.64***	***83.9 ± 8.64***	***83.0 ± 9.06***	***81.8 ± 8.59***

° Plus-minus values are means ± SD. Percentages may not total 100 due to rounding.

Ɨ Missing values were recorded using imputation.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 2. Logistic Regression Model Output.

Coefficients	**Model 1** *(Multiple Times v. Once)*				**Model 2** *(Once v. Never)*				**Model 3** *(Multiple Times v. Never)*			
	Estimate	Std. Error	*Z* Value	*p* Value	Estimate	Std. Error	*Z* Value	*p* Value	Estimate	Std. Error	Z Value	*p* Value
Intercept	-1.668	0.869	-1.919	0.055	-3.754	1.274	-2.948	0.003**	2.181	1.519	1.435	0.151
Gender (Male)	-0.713	0.165	-4.318	0.00***	-0.654	0.245	-2.666	0.008**	-0.015	0.293	-0.050	0.960
Race (Black)	-0.167	0.307	-0.544	0.586	-0.129	0.410	-0.314	0.754	-0.153	0.517	-0.296	0.768
Race (Hispanic)	-0.772	0.359	-2.153	0.030**	-1.373	0.585	-2.346	0.019*	0.631	0.696	0.907	0.364
Race (Native American)	-0.672	0.770	-0.872	0.383	-0.721	1.068	-0.675	0.500	0.042	1.330	0.032	0.975
Race (Other)	-0.387	0.647	-0.598	0.550	-0.130	0.790	-0.165	0.869	-0.030	1.016	-0.029	0.977
Race (White)	-0.467	0.240	-1.949	0.050*	-0.706	0.315	-2.240	0.025*	0.162	0.408	0.397	0.691
SAT Math	0.002	0.002	0.976	0.329	0.004	0.002	1.513	0.130	-0.001	0.003	-0.422	0.673
SAT Verbal	-0.001	0.002	-0.152	0.880	-0.001	0.002	-0.629	0.530	0.001	0.003	0.311	0.756
SAT Writing	-0.002	0.002	-1.136	0.256	-0.001	0.002	-0.272	0.786	-0.001	0.003	-0.510	0.610
ACT Composition	-0.071	0.105	-0.674	0.500	-0.009	0.153	-0.060	0.952	-0.005	0.191	-0.025	0.980
ACT English	0.021	0.044	0.488	0.626	-0.104	0.065	-1.598	0.110	0.120	0.078	1.532	0.126
ACT Math	0.033	0.050	0.654	0.513	0.048	0.073	0.665	0.506	-0.057	0.092	-0.620	0.535
ACT Reading	0.023	0.042	0.544	0.587	0.049	0.061	0.802	0.422	-0.053	0.076	-0.697	0.486
ACT Writing	0.003	0.039	0.084	0.933	0.030	0.058	0.517	0.605	-0.020	0.075	-0.269	0.788

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 3. Unmatched Observations by Treatment.

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

Table 4. Estimating Treatment Effects.

Treatment	Mean Difference	Confidence Interval	p value
Multiple Visits – Never	+ 2.061	(1.646, 2.476)	$3.888 \times 10^{-22***}$
Multiple Visits - Once	+ 1.619	(1.206, 2.031)	$1.894 \times 10^{-14***}$
Once - Never	+ 0.443	(0.019, 0.875)	$4.486 \times 10^{-2*}$

$^{*}\ p < 0.05$; $^{**}\ p < 0.01$; $^{***}\ p < 0.001$

APPENDIX B

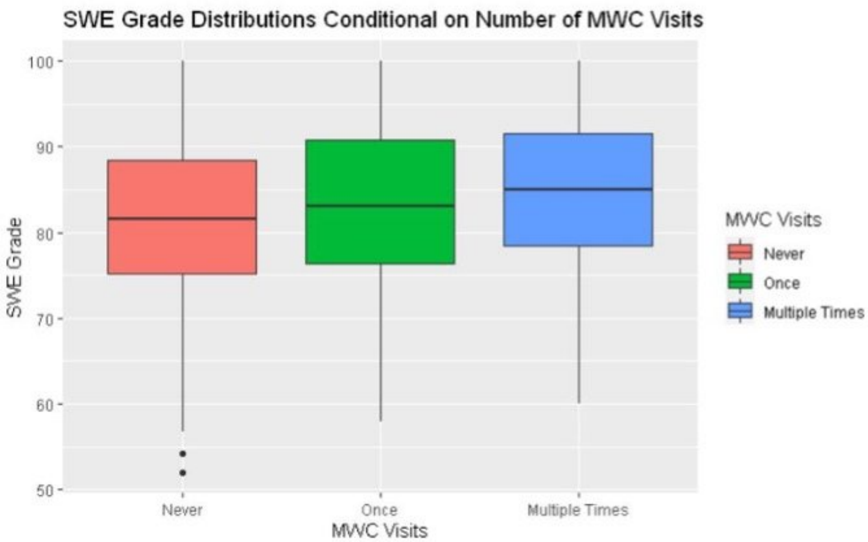
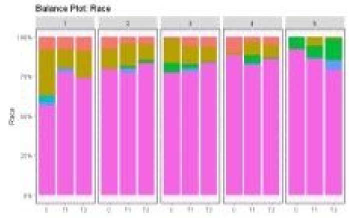
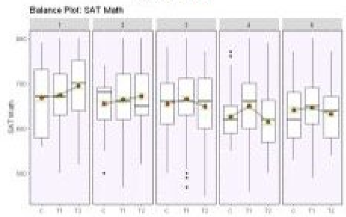


Figure 1: SWE Grade by Number of MWC Visits

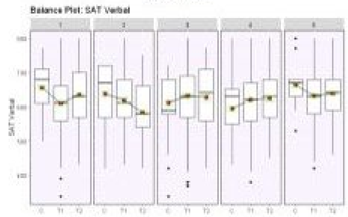
(a) Gender



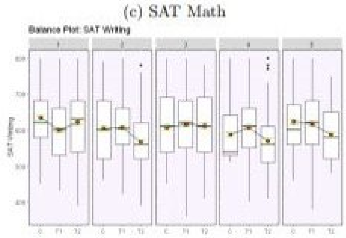
(b) Race



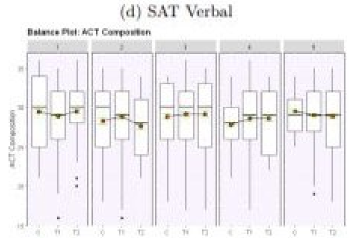
(c) SAT Math



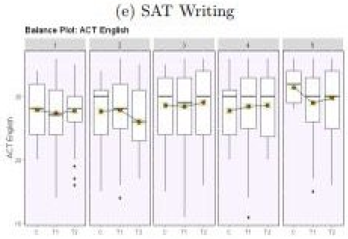
(d) SAT Verbal



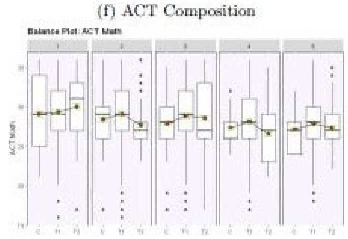
(e) SAT Writing



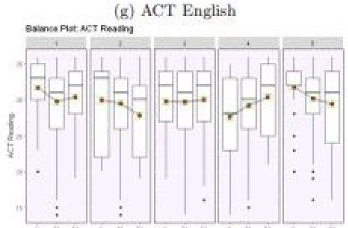
(f) ACT Composition



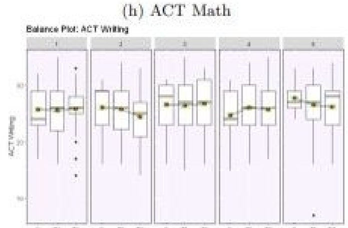
(g) ACT English



(h) ACT Math



(i) ACT Reading



(j) ACT Writing

Figure 2: Covariate Balance Plots

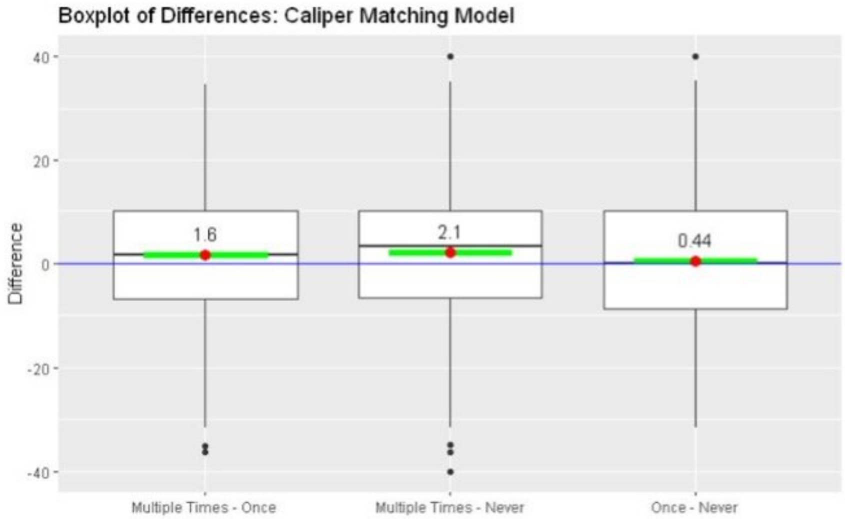


Figure 3: Mean Difference in SWE Grade Between Treatment Groups