

1-1-2016

"By Turns Pleased and Confounded": A Report on One Writing Center's RAD Assessments

Scott Pleasant

Luke Niiler

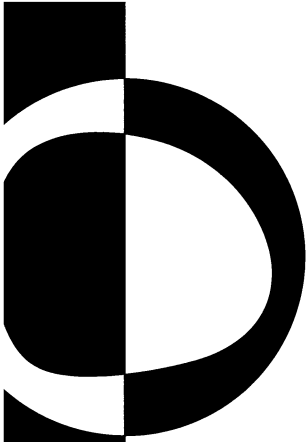
Keshav Jagannathan

Follow this and additional works at: <https://docs.lib.purdue.edu/wcj>

Recommended Citation

Pleasant, Scott; Niiler, Luke; and Jagannathan, Keshav (2016) ""By Turns Pleased and Confounded": A Report on One Writing Center's RAD Assessments," *Writing Center Journal*: Vol. 35 : Iss. 3, Article 6. DOI: <https://doi.org/10.7771/2832-9414.1843>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.



Scott Pleasant, Luke Niiler, & Keshav Jagannathan

“By Turns Pleased and Confounded”: A Report on One Writing Center’s RAD Assessments

Abstract

This study builds on extant replicable, aggregable, and data-supported (RAD) research to posit and examine correlations between writing center intervention and improved student writing. The authors review three decades of quantitative writing center scholarship and provide data resulting from four writing center assessments. These assessments include two pre- and post-intervention studies and two intervention/non-intervention studies. Results are mixed. The pre- and post-intervention studies show statistically significant improvements in student writing. The intervention/non-intervention studies show considerable to limited improvements in student writing. Possible reasons for these results are discussed, including study protocols, self-selection bias, and the difficulty of imposing controls. Impacts on the practice of one writing center are shared, and suggestions for further research are provided.

The Writing Center Journal 35.3 | 2016 105

At a time when academic programs are increasingly evaluated by quantitative measures, more writing centers should consider the benefits of assessments driven in part by replicable, aggregable, and data-supported (RAD) research methods. As our study demonstrates, such assessments can productively interrogate and explicate the myriad roles tutoring plays in the improvement of student writing. This idea is not new. Harvey Kail (1979) writes, “Even if peer tutoring is a relatively new educational innovation and even if it may be still premature to begin a systematic evaluation of its effectiveness, my Dean, at least, wants to know (and sooner, not later), ‘Does your peer tutoring program work?’” (p. 2).

Kail’s call has remained largely (although not entirely) unanswered, though, and for some understandable reasons. In “Writing Center Assessment: Why and a Little How,” Isabelle Thompson (2006) suggests that some directors are wary of implementing quantitative measures of success because they “are already overwhelmed with other duties.” Indeed, writing center administrators may “equate externally-mandated assessment with external accountability to conservative institutions not particularly supportive of our process-based pedagogy” (p. 33). Others may view quantitative methods skeptically, believe that quantitative methods are beyond their purview, or fear the potential results of quantitative assessments. If a center is judged by its performance on such measures, what happens when it performs poorly? Is its budget cut? Is it shut down entirely?

While these concerns are certainly valid, it is our view that the advantages of articulating and implementing robust quantitative assessment protocols outweigh these risks. As we show below, if in fact RAD writing center research occasionally confounds our expectations, those expectations can be adjusted. While RAD research may not always affirm our work, it can provide a clearer view of our work. And that view can become a vision of how to do that work better.

Recently, calls for quantitative studies of writing center work and advice on how to conduct such studies have intensified. Ellen Schendel & William J. Macauley’s *Building Writing Center Assessments that Matter* (2012) includes a chapter by Macauley titled “Getting from Values to Assessable Outcomes” along with an “interchapter” by Neal Lerner titled “Of Numbers and Stories: Quantitative and Qualitative Assessments in the Writing Center.” The underlying message from both Macauley and Lerner is that we can and should quantify writing center effectiveness. Macauley focuses on turning our values into “assessable outcomes” and identifies broad goals such as fostering close reading and critical inquiry skills among students (p. 41) and encouraging tutors to

use non-directive methods and focus on higher-order concerns before lower-order concerns (p. 44). Macauley offers three questions that must be asked in determining whether an outcome is assessable: 1. "Can it be measured or counted?" 2. "Can it be measured or counted consistently?" and 3. "Is it clearly a reflection of the value to which you have attached it?" (p. 52). This kind of language—borrowed in equal measures from the natural and social sciences and from education—suggests that more could be done to augment the qualitative and anecdotal measures that have traditionally dominated writing center research.

The title of Lerner's interchapter also indicates the value he places on both qualitative and quantitative measures. Lerner describes the "conflict between numerical data versus qualitative data or numbers vs. stories" not in simple terms that portray one as "better" than the other but as a "conflict in the fundamentals of knowledge making" (p. 109), and he argues that both are essential if we want to paint a clear picture of our work. Lerner asks his readers to imagine a new dean asking for a report on the writing center's effectiveness. The director is torn between providing neat tables of numerical data or gathering testimonials from students and tutors. In the end, both are valuable, he says, because

[n]ew deans do not always demand bar charts and summary tables for you to demonstrate the value of your writing center. Some might be more persuaded by narrative accounts of how student writers learn or a description of the long-term value that your tutoring staff finds from their work. (p.113)

In *Researching the Writing Center: Towards an Evidence-Based Practice* (2012), Rebecca Day Babcock & Terese Thonus announce, "The purpose of this book is first to argue for RAD research, qualitative and quantitative scholarship that engages empirical evidence as mediating theory and practice" (p. 3). They trace a history of scholarship that has been uneven at best in its appreciation for quantitative work. The value of combining quantitative and qualitative data seems evident, but calls for quantitative work to complement our "stories" are, for the most part, a recent phenomenon. It's not that nobody has *ever* said writing center research should incorporate more quantitative data, but few were saying it loudly and even fewer were publishing truly quantitative work. In "Theory, Lore, and More: An Analysis of RAD Research in *The Writing Center Journal*, 1980–2009," Dana Driscoll & Sherry Wynn Perdue (2012) look at thirty years' worth of articles, 270 in all, and conclude that only 6% of those articles—a total of 15—would

qualify as RAD research (p. 25).¹ Driscoll & Perdue’s analysis echoes in many ways Richard Haswell’s (2005) plea for more RAD research in composition. In “NCTE/CCCC’s Recent War on Scholarship,” he shows that composition researchers published more RAD research than did their counterparts in the writing center field. Haswell identifies 69 RAD articles on research paper writing from the 1990s in non-NCTE/CCCC publications, but only three in NCTE/CCCC publications. Perhaps this dearth of RAD research has something to do with the perceived consequences of “negative” findings—that is, findings that would seem to contradict the efficacy of the writing center. Perhaps, too, as James Williams & Seiji Takaku (2011) note, many writing center researchers labor under the belief that there are simply too many variables to produce a replicable, aggregable, and data-supported study. There is, after all, considerable variation in teaching and grading “from teacher to teacher, from class to class, and from tutor to tutor” (p. 6). So we have two issues in play as we consider the role of RAD research in our field: First, the consequences of findings that do not meet our expectations, and may not meet the expectations of our institutions; and second, perceived difficulties with designing a reliable RAD study that attempts to measure the efficacy of writing center tutoring.

We will touch on both of these issues in this article, which focuses on a series of RAD assessments over two academic years. These assessments provide not only what we hope will be a model, but also encouragement, caution, and sometimes puzzling data on the effectiveness of peer tutoring. Our RAD assessments employ quantitative methods to demonstrate that writing center tutoring may lead to better student writing. Perhaps just as importantly, though, they have helped us better understand our current practices with an eye to improving them. And they have encouraged us to look for new and more innovative ways to employ quantitative methods within the context of RAD research in future assessments.

Selected Literature Review

The roots of RAD writing center research, however thin, stretch deep. Kail’s 1979 call, cited above (“Does your tutoring program work?”), seems to echo throughout Stephen North’s 1984 essay, “Writing Center

1 While only a very small number of articles qualify under the authors’ rigorous standards for RAD research, Driscoll & Perdue did find that 91 of those articles contained some form of data collection, either first-hand collection from human participants or from secondary sources (p. 20).

Research: Testing Our Assumptions." In that work, North calls for research methods "borrowed from disciplines like ethnography, social psychology, and cognitive psychology." North advocates videotaping, transcription, and protocol analysis as means of identifying effective tutoring practices (p. 30). For many years after, quantitative writing center researchers were focused primarily on grades and retention. Examples include Mark Waldo (1987), who finds positive correlations between writing center intervention and grades. David Roberts' 1988 essay, "A Study of Writing Center Effectiveness," describes a pre-test/post-test design to compare classroom writing instruction with individualized writing instruction on two college campuses. Roberts finds no significant differences. Ellen Mohr's 1995 study, "Researching the Effectiveness of a Writing Center," shows statistically significant differences between the course grades of an "experimental" group of students who were encouraged to visit the writing center and those of a "control" group who were not. Neal Lerner's "Counting Beans and Making Beans Count" (1997) is his first of many passes at quantitative assessment. Like Mohr, Lerner demonstrates that students who receive tutoring earned higher grades than students who did not and that students with the lowest SAT scores benefitted the most from tutoring. Stephen Newmann (1999) echoes Lerner's findings in his study, "Demonstrating Effectiveness," in which he shows that tutoring helps "less able students [as determined by SAT scores] who were willing to work harder perform as well as their peers" (p. 9).

More recently, Gary Griswold (2003) and Diana Calhoun Bell & Alanna Frost (2012) examine the relationship between writing center use and student retention. The researchers find that retention rates for writing center users are higher than for those of non-users, but they also note that it is difficult to control for self-selection bias in the use of writing centers. That is, students who use writing centers may be more motivated than non-users, and therefore more persistent. Similarly, Katherine M. Schmidt & Joel E. Alexander (2012) find higher rates of "writerly self-efficacy," students' belief in their own ability to successfully complete a task, among writing center users than non-users. Julia Bredtmann, Carsten Crede, & Sebastian Otten (2013) find that writing center intervention has no significant impact on student writing as measured by student performance on end-of-term writing exams. (Curiously, the researchers chose not to examine the actual work performed by writing center consultants, who did not help students prepare for writing exams but did help with drafts of theses and dissertations.)

These noteworthy examples notwithstanding, quantitative writing center research has generally avoided the issue of the effect of

writing center tutoring on student writing. As Thompson (2006) notes, many of these studies have not addressed the core question of *whether the writing center actually helps students write better*. For example, Thompson recalls that during a year-long program in which her center expanded its services, “The best way I knew to justify our existence was to collect usage data and conduct student and faculty satisfaction surveys” (p. 35). Thompson’s well-intentioned but ultimately inconclusive practices have been replicated in center after center. Casey Jones (2001), in turn, concludes that an “exhaustive search of the literature reveals that only a handful of researchers have attempted to evaluate the performance of writing centers in enhancing student writing skills through the use of empirical study designs” (p. 3).

However, a few more recent studies more directly consider the role of the writing center in improving student writing. These include the work of Williams & Takaku (2011), who review eight years of student data, including writing exams, standardized test scores, and course grades. The researchers note that students who more frequently seek writing center assistance receive higher composition grades than those who do not. Roberta Henson & Sharon Stephenson (2009) attempt to learn if writing center intervention improves overall writing skills during the course of one semester. They compare writing center clients’ and non-clients’ first and last essays from one semester, and conclude that writing center clients’ work shows statistically significant improvements in several areas (thesis, examples, sensory detail, paragraph unity, and overall quality), while non-clients’ work does not.² More recently, L. Lennie Irvin (2014) finds that three seems to be a magic number: That is, he determines that significant improvements in several student writing attributes occur, in general, after three tutorials. Irvin sees improvement in student persistence, retention, and GPA. Interestingly, Irvin also sees an uptick in professors’ “PGR,” or “progressive grade rate,” a calculation of the number of “C” or better grades they assigned prior to and during the study. Heather Robinson (2009) shows that after three tutorials, writing center clients demonstrate increased intrinsic motivation—that is, they find a personal reward in visiting the writing center of their own free will, and not at the behest of a third party. A 2010 study by Rowena Yeats, Peter Reddy, Anne Wheeler, Carl Senior, & John Murray demonstrates that the number three again has signifi-

2 It should be noted that Henson & Stephenson reviewed essays “without regard to whether students had been to the writing center” for those specific [i.e., first and last] essays. Their interest was on skills development, not evidence of writing center impact on specific papers.

cance: this time, in terms of higher final grades, as the work of writing center attendees is compared to that of non-attendees.³

Our current work draws from and builds upon two studies. Nancy McCracken (1979) details a pre-test/post-test evaluation protocol focused on error elimination. In her study, students provide the writing center with writing samples at the start of each semester, and writing center staff identify errors in them. After being tutored on these errors for one semester, students provide an additional writing sample. Errors are again identified, and the frequency of error in the first and second drafts is compared. Luke Niiler (2003, 2005) employs a multiple-trait analysis protocol to compare pre- and post-intervention versions of the same papers. In both studies, post-intervention drafts showed statistically significant improvement over pre-intervention versions in each trait measured.

Two Years of Assessments

The present study—or series of related studies—seeks not only to enter into the arena of quantitative RAD research but to expand on Niiler’s studies. Building on Niiler’s pre-intervention/post-intervention protocol, our study provides an example of the kind of RAD research that can be conducted. We will describe in detail how the plan was designed and implemented and give some encouragingly positive data along with some puzzling data that will, we hope, spur others to replicate and extend the work we have done. We hope that this kind of research can prove instrumental in not only enriching the research profile of our field, but also in augmenting our daily practices, and, perhaps, offering you a template for your own RAD assessments.

In the year before beginning a series of quantitative assessments, the writing center at Coastal Carolina University followed the common approach of counting annual visits and comparing that data to totals from earlier years. In addition to documenting a 70% increase in tutorials over the previous year, the center’s annual report included survey results showing that nearly 100% of students “agreed” or “strongly agreed” with statements like “I would recommend this service to others” and “The tutor was helpful.” That data demonstrated that students were satisfied with the writing center, but did not demonstrate how the writing center actually impacted student writing. Therefore, after one year of

3 Indeed, Irvin’s, Robinson’s, and Yeats, Reddy, Wheeler, Senior, & Murray’s work would seem to confirm a central tenet of Williams & Takaku’s work (2011): Student success in writing centers can be linked to “help-seeking behavior” (p. 6).

this approach, we expanded our assessments to include data that would attempt to quantify if, and how much, students' writing improved after visiting the center.

Near the end of the 2010–2011 academic year, the Coastal Carolina University writing center applied for an internal assessment grant that provided the opportunity to design the kind of quantitative study that Jones, Thompson, Niiler and others have called for. Before this study, we routinely collected data on various indirect measures of effectiveness—student satisfaction surveys, number of one-with-one visits, attendance at workshops, number of tutors at training sessions, faculty opinions, and even results from written tests taken by tutors. However, these kinds of data do not provide direct, observable measures of how, and how much, writing center tutoring impacts the quality of student writing.

Direct assessments come in two basic varieties: pre-intervention vs. post-intervention studies and intervention group vs. non-intervention group studies. Pre- vs. post-intervention studies consider drafts written before students visit the writing center and compare them to revised versions after students have visited the writing center. Intervention group vs. non-intervention group studies compare papers written by students who visited the writing center at some point during the writing process to papers written by students who did not. It should be noted that the terms “control group” and “test group” are not entirely accurate for this kind of assessment because they imply that the researchers have eliminated all other possible causal factors for students' writing improvement other than writing center tutoring. A study of this type can never eliminate all other possible factors—nor would any responsible researcher wish to do so. Therefore, in a strict sense, studies that group papers into a “test group” and a “control group” are more accurately called “intervention” vs. “non-intervention” studies. In the first semester of the study, we chose the pre- vs. post-intervention model. Then, in the second and third semesters, we added intervention vs. non-intervention group protocols.

Our Model: Niiler's Pre-/Post-Intervention Studies

As stated above, the model we employed for our assessment study was Niiler's work from his 2003 and 2005 studies. Those studies address the question we were most concerned with: Does a student's writing improve as a result of a writing center tutorial, and if so, how much and in what areas? While Niiler's two studies represent an important leap forward in quantitative writing center assessment, they have sev-

eral limitations, so we attempted not to replicate his valuable work but augment his approach.

In a 2003 article Niiler encourages such attempts. He emphasizes that his first assessment is merely a "pre-test, a means of not only finding answers...but also a way of creating a better test" (p. 6). For example, he admits that because he "did not employ an adequate control, I have to place a large asterisk next to these figures" because "I cannot unequivocally claim that the writing center actually 'caused' improvement in the writing of those who visited with us" (p. 8). Niiler is correct in noting that a lack of true control data (which is difficult, if not impossible, to include in such as study) weakens any claim of causality.

Our first study attempts to deal with some limitations in Niiler's research designs. In the 2003 study, the raters (tutors from the writing center) knew whether they were rating a pre-intervention or post-intervention version of any given paper, a fact that could have biased their ratings. "Blind" rating of drafts—ratings performed by readers who don't know whether any given paper is a pre- or post-intervention version—avoid this potential for bias.

Niiler recognizes the lack of "blindness" in his first study and asks, "What of the possibility of rater bias, given that the raters were, in fact, tutors?" Reflecting now on this matter, Niiler feels that raters may have assigned higher ratings in order to favorably influence his perceptions of their tutoring ability. When this potential for bias is added to the bias inherent in tutors rating the effectiveness of their own work, there is reason to question the validity of Niiler's results in the 2003 study. In the follow-up study published in 2005, he addressed these potential biases by seeking raters from outside the writing center and putting in safeguards to prevent raters from knowing whether any given paper was a pre- or post-intervention version. Perhaps not too surprisingly, the mean improvement score decreased somewhat from .7 points on a five-point scale in the first study to .57 points in the second study.

Potential problems with the study do not end with the issue of the raters' biases, however. One issue in both studies is that no attempt was made to gather similar kinds of papers (i.e., responses to similar writing prompts or similar kinds of writing tasks) or to collect papers from students at the same level. Both studies mix papers from upper-division courses with those from lower-division courses in a variety of disciplines and in response to a variety of writing assignments. (Niiler now notes that this mix of papers was by design, as he was attempting to demonstrate the efficacy of tutoring across disciplines). Also, Niiler admits in his 2005 article that raters in his studies—especially the three

faculty members in the second study—have a “need for stronger rater training” (p. 14).

The consistently positive results in Niiler’s studies should be read with some caution, or in Niiler’s own words from the 2003 study, “a large asterisk” (p. 8). Still, these two empirical studies put his work ahead of a coming trend. The studies described below represent attempts to follow and improve upon Niiler’s model.

Writing Center Assessment: Year One

Methodology. During Spring 2011, the Coastal Carolina University writing center received an internal grant for assessment in the 2011–2012 academic year. In the funding application, we proposed collecting pre- and post-intervention drafts of research papers written in first-year courses. These papers would be gathered from students who visited the writing center and agreed to participate in the assessment. After one “intervention” (i.e., a standard 30-minute writing center tutorial), these students were to return with completed post-intervention revisions of the same papers. At the end of the semester, the papers we collected would be rated “blindly” by a panel of six readers from the English Department faculty. All of the raters were seasoned full-time instructors or professors who taught multiple first-year composition courses during that academic year and previous years. No rater would read both the pre- and post-intervention version of the same paper, and in a further attempt to mitigate rater bias, each paper would be rated by two different readers and those scores would be averaged. We would then calculate the average ratings for both pre- and post-intervention papers and conduct statistical analyses on the results.

The assessment plan called for the collection of research papers from three first-year courses: English 101, University 110 (a first-year seminar), and Physics 103 (a general science course that includes a research paper requirement). In order to be deemed eligible for the study, papers had to incorporate multiple outside sources and be formatted according to MLA or APA style. We chose these criteria for three reasons. First, the writing center often provides help with these sorts of assignments, so it seemed logical to assess our effectiveness with one of our more common tasks. Second, these types of papers give us a chance to evaluate our success with a wide variety of issues: integration of sources, organization, thesis and paragraph development, sentence structure, and formatting. Third, we felt that we should evaluate our effectiveness with an assignment that is common across the disciplines.

At the end of the Fall 2011 semester, 40 pre-/post-intervention paper pairs (80 papers total) were rated both holistically and according to a specific set of six criteria (described later in this article) by a panel of six faculty readers trained and monitored by a "table leader" with several years' experience working with the AP holistic grading system. We chose to do both holistic and criteria-based ratings because we felt that both kinds of ratings were necessary in order to understand the differences between the pre- and post-intervention papers. The holistic rating would tell us whether the post-intervention papers had, in an overall sense, improved. The criteria ratings would tell us exactly where the papers had improved and where they had not. By looking at both holistic ratings and ratings for specific criteria, we could, at least tentatively, conclude that average improvements seen in the holistic scores were attributable to, say, improvements in the thesis score or the development score. The criteria ratings would also help us begin to understand what areas our tutors were doing their best work in and where they might need some additional training.

The day-long process of rating the papers began with a three-hour "norming" during which the table leader helped the panel of raters achieve consistency in their ratings. As part of that process, he distributed "anchor" papers representing each of the nine levels on a nine-point holistic scale. These papers were gathered from previous assessments done by the English Department, not from papers in the writing center study. The raters read these benchmark papers and reached consensus on why and how each paper represented a particular score.

The panel then rated the papers in the study during an afternoon session. The holistic score ranged from 1–9, with 1 the lowest and 9 the highest rating. In consultation with the table leader, we developed score descriptions based on the nine-point AP grading system but tailored to writing expectations in our first-year writing program. Table 1, below, gives the detailed descriptions used in the study for each of the nine holistic levels:

Table 1: Descriptions of Holistic Levels, 2011–2012

Points	Description
9	This paper goes significantly beyond the expectations for college-level writing. Writing at this level is fully mature in thought, organization, and development. The writing style in this paper is at a professional level, and there are very few, if any, errors in grammar/punctuation/usage/spelling. Formatting and documentation are near perfect.
8	This paper is characterized by many or most of the elements of maturity in the 9-point paper, but there are some minor weaknesses in one or more areas. On the whole, though, this writing is well above the average for college-level writers in all areas.
7	This paper represents solid college-level work but demonstrates some weakness or lack of maturity in the areas of thought, organization, or development and/or may include noticeable problems with or errors in style, grammar/punctuation/spelling/usage or formatting/documentation.
6	Like the 7-point paper, this paper is good college-level work, but the 6-point paper will demonstrate weaknesses indicative of the developing writer. Thought, organization, or development may lack maturity or be somewhat ineffective or insufficient. Grammar/punctuation/spelling/usage may feature noticeable, even distracting, errors. Formatting and documentation may need improvement. This paper is slightly above average at the college level.
5	Overall, this paper represents average college-level writing that does not rise above the average for developing writers at this level. Weaknesses or lack of maturity in thought, organization, and development are common but not especially problematic. Grammar/punctuation/spelling/usage errors may be found throughout but do not indicate an especially weak ability to control the elements of Edited American English. Formatting and documentation will often need to be polished and improved but in most cases will represent a college-level attempt to follow such guidelines.

Points	Description
4	This paper is similar to the 5-point paper in that it is indicative of college-level writing ability, but noticeable weaknesses at one or more levels mark this paper overall as slightly below the average for developing writers at the college level.
3	A paper at the 3-point level does not represent college-level work but does demonstrate some control of and skill with organization and development. Thought may be significantly lacking in maturity at this level, and errors in or problems with style and grammar/punctuation/spelling/usage are often more noticeable or more severe than in the average paper. Formatting and documentation may be especially problematic, but these errors alone should not mark a paper as a 3.
2	A paper at the 2-point level features many of the same weaknesses as the 3-point paper, but at this level, the problems or errors are more severe or more common. Still, this paper demonstrates some control of the standards of Edited American English and represents a bona fide attempt to produce a college-level paper even though the paper falls well below college-level standards.
1	This paper is problematic throughout at all levels. This paper not only represents writing below the college-level but is nearly unreadable due to numerous distracting errors in grammar/punctuation/spelling/usage. Thought is not just immature, but is expressed in such confusing or disorganized language as to be nearly impossible to decipher. In short, this paper is so far below the expectations for college-level writers that it gives very little indication of the writer's being prepared to make the improvements necessary to write at the college level.

While no rubric can completely capture the complexities of writing, we feel that the descriptions above provided the raters with a relatively clear set of standards for the purposes of the study.

In addition to the holistic score, each paper would be rated on a five-point scale in each of six categories: thesis, organization, development, style, surface, and presentation. In consultation with the table leader, we decided on a five-point scale for the specific categories

(similar in most respects to Niiler’s “traits”) because we felt that it would be difficult to justify nine levels of sensitivity. While it might be relatively easy to distinguish between a seven- and an eight-point paper on a holistic scale, we felt it would be difficult for most readers to make similarly fine distinctions on a category like style or organization.

The first three categories cover higher-order concerns. The thesis category rates the writer’s ability to state and maintain a clear thesis throughout the paper. In the organization category, raters looked at the writer’s ability to put ideas in a logical order and make connections between those ideas. The development category rates the amount and quality of development in a paper. This category recognizes that developing ideas with specific, relevant content is separate from the ability to state a thesis or organize ideas within a paper.

The next three categories comprise lower-order concerns. The style category refers to sentence-level and paragraph-level issues such as appropriate tone, sentence variety, and skillful integration of paraphrased and quoted source material. The surface-level category involves adherence to conventions of grammar, usage, and punctuation. The presentation category evaluates the writer’s ability to conform to APA or MLA formatting guidelines.

Statistical analysis. In this section, we will define the three main statistical concepts used in this article in lay terms, which should help clarify our results and discussion.⁴ The data we show below involves *means, statistical significance, p-values, and t-tests*. Means are numerical averages, which are used in this article to report raters’ findings. Statistical significance indicates the probability that an occurrence can be attributed to cause, not chance. That is, we see statistically significant results if the results of a statistical test cannot be attributed to chance. To measure significance, we have the p-value. P-values measure the possibility that the observed result is solely due to chance. When p-values are smaller than an established level (.05, or five percent, in most statistical literature), the results are statistically significant. *We reach the level of statistical significance when our results are shown to be at least 95% likely to be due to writing center intervention rather than chance alone.* As you review our data in the tables below, look closely at the p-values in the

4 Statistical techniques are divided broadly into two categories: *parametric* and *non-parametric*. A parametric technique is one that assumes that the data follow a curve (typically, the bell curve) and performs analyses under that assumption. Because our data do not follow a bell curve, we are using non-parametric techniques—specifically, a Wilcoxon test, which compares two related samples. Only those results immediately relevant to this report are shown here. For complete statistical data, please contact the authors.

far right-hand column. If you see a p-value of 5% (.05) or less, you can assume the results are statistically significant. It is probable that those results can be attributed to cause—that is, writing center intervention. If the p-value is greater than .05, you will know that the results are not statistically significant, and the results are probably due to chance. The significance of our statistical data is further clarified in our “Results” and “Discussion” sections.

P-values are calculated by t-tests, or the appropriate statistical technique for studies such as ours in which differences in means (in this case, the averages of raters’ ratings of student writing samples) are compared. T-tests can be either one-sided or two-sided. One-sided t-tests are used when one kind of result is expected. That result might be improvement or deterioration. Two-sided t-tests are used when results are not necessarily expected to be either better or worse. In this article, we employ two-sided t-tests. Even though we might expect only one kind of result (improvement) from writing center intervention, we cannot know this for certain. A two-sided t-test allowed us to avoid making that assumption; we wanted to leave open the possibility that a visit to the writing center might not necessarily result in an improved paper.

Year one results and discussion. After the panel completed rating, the results were encouraging. On the holistic scale, the 40 draft papers averaged 3.7 on the nine-point scale while the revised papers averaged 4.25. The average improvement was 14.9%, for an average rise of .55 points. Put another way, the 3.7 average for the drafts equals 41.1% of the nine possible points on this measure. The 4.25 average on the revisions equates to 47.2%, giving a percentage-point improvement of 6.1% or, in terms of letter grades, slightly over half a letter grade.

Scores on the six specific categories are similarly positive. Table 2, below, shows how raters rated drafts that had not been reviewed in the writing center (as indicated by the ‘Pre Mean’ column), and how they rated those same drafts after writing center review (the ‘Post Mean’ column). The next column, ‘Raw Diff,’ indicates the difference between the values shown in the ‘Pre Mean’ and ‘Post Mean’ numbers. The fourth column, ‘% Diff,’ is a calculation of the percentage difference between ‘Pre Mean’ and ‘Post Mean’ numbers.

Table 2

Multiple Trait Assessment Data, Fall 2011

Category	Pre Mean	Post Mean	Raw Diff.	% Diff.
Thesis	2.44	2.71	.27	11.3%
Organization	2.45	2.70	.25	10.2%
Development	2.40	2.74	.34	14.1%
Style	2.51	2.66	.15	6.0%
Surface	2.53	2.76	.23	9.4%
Presentation	2.75	3.10	.35	12.7%
Mean (n=40)	2.51	2.78	.27	10.6%

For example, the difference between the averages of pre- to post-intervention in the “Thesis” category is .27; this represents an 11.3% increase from pre- to post-intervention. When all of the category ratings are combined, the rating among the Fall 2011 papers went from 2.51 on the pre-intervention papers to 2.78 on the post-intervention papers for a mean improvement of .27 points on a five-point scale or 10.6%.

This amount of improvement seems significant, but our statistical analysis featuring the two-sided t-test—which is, again, the test we use to determine p-values when we can’t predict if results will show improvements or deterioration—demonstrates statistically significant results in the “Development” and “Mean” categories only.⁵ See Table 3, below.⁶

5 Please Note that ‘Holistic’ ratings are included in Tables 3, 5, and 11 for reference. We do not include ‘Holistic’ ratings in ‘Mean’ scores, as they represent a separate rating methodology.

6 Table 3 can be read in much the same manner as Table 2, with ‘pre’ and ‘post’ designating ratings of student drafts prior to and after writing center intervention.

Table 3*Statistical Analysis, Fall 2011*

Category	Pre Mean	Post Mean	p-value (two-sided)
Holistic	3.705	4.282	0.0680
Thesis	2.438	2.713	0.0920
Organization	2.45	2.70	0.1390
Development	2.40	2.738	0.0300
Style	2.513	2.663	0.1948
Surface	2.525	2.7625	0.1100
Presentation	2.75	3.10	0.0618
Mean (excluding Holistic) (n=40)	2.513	2.7792	0.0362

This is to say: The only categories that we can report as having benefited from writing center intervention are “Development” and “Mean” (“Mean” being the average of all categories rated). No other category demonstrated statistically significant results. As you’ll note, Table 3 shows that the “Holistic,” “Presentation,” and “Thesis” categories show p-values under .10, an indication that these categories approach, but do not meet, the .05 standard of statistical significance. “Style,” “Surface,” and “Organization” neither approach nor meet the .05 standard. These p-values are very important, because they demonstrate that any improvements in these six categories are likely due to chance. P-values must be at .05 or less for us to claim that improvements were not due to chance.

With this data, we can posit a link between writing center tutoring and improvement in “Development” and “Mean” scores. However, as this first study is strictly a pre versus post type and lacks control data, it cannot establish with certainty that writing center intervention is the major cause of the kinds of limited improvements seen in the other categories noted above. The students involved in the study also had instruction by their classroom teachers, peer reviews in class, and (potentially) other factors contributing to their improvement. Also, as with many studies, this one does not definitively separate correlation

and causation. Thus, it cannot establish that those students' scores would not have improved without this intervention. Our study cannot definitively disprove the "null hypothesis," which would conclude there is no clearly-established relationship between the pre- and post-intervention data.

There is also the possibility of two kinds of self-selection bias in this study. First, it is possible that students who voluntarily come to the center and agree to participate in a study are simply better or more motivated students than average. Their papers could therefore be expected to improve. It is also possible that better or more motivated students are more likely to contribute not only pre-intervention drafts at their writing center conferences, but also return to drop off their revised papers at a later date. We did not rate the papers of students who contributed only a pre-intervention version and never provided a post-intervention version. If those unrated pre-intervention drafts had been rated lower than the rated pre-intervention drafts, more improvement might have been shown. All of these limitations and qualifications exist, and perhaps help explain, in part, why there has been the dearth of RAD research noted above. Yet we have taken a small, but encouraging step with this study. The positive data suggests that there is a stronger correlation between writing center intervention and writing improvement in terms of "Development," even though there is a much weaker correlation between intervention and improvement in terms of "Presentation," "Thesis," and "Holistic" scores.

During the Spring 2012 semester, we conducted a second round of pre-/post-intervention ratings of research papers in first-year writing classes. Aware of some limitations in the study design as described above, we wanted to modify our methodology. However, we maintained virtually the same methodology from Fall 2011, as our IRB approval had been granted for one year under those terms. In Spring 2012, only one rater scored each paper (whereas two raters had scored each paper in Fall 2011). Had we had maintained the standard of two raters for each paper, the six-member panel may not have had time to rate papers during the one day allotted for the study.

The results of the Spring 2012 study were similar to the Fall 2011 results. A strong push for recruitment of more papers resulted in 62 pairs, as compared to 40 pairs in Fall 2011. On the nine-point holistic scale, the pre-intervention papers averaged 4.21 points, while the post-intervention papers averaged 4.79 points. Average improvement was .58 points in Spring 2012 while it had been .55 points in Fall 2011.

On the six specific categories, again the results from Spring 2012 were similar to the results from Fall 2011, as shown in Table 4⁷, below:

Table 4

Multiple Trait Assessment Data, Spring 2012

Category	Pre Mean	Post Mean	Raw Diff.	% Diff.
Thesis	2.60	2.97	.37	13.8%
Organization	2.61	3.02	.40	15.4%
Development	2.66	2.97	.31	11.5%
Style	2.74	3.05	.31	11.2%
Surface	2.87	3.02	.15	5.1%
Presentation	2.68	3.19	.52	19.3%
Mean (n=62)	2.69	3.03	.34	12.7%

The average overall improvement was .27 points (or 10.6%) on a five-point scale in 2011. The mean, or average, in Spring 2012 was .34 points (or 12.7%). In each category other than "Surface," the results show an average improvement of between 10 and 20 percent. The ratings for "Presentation" show not only the highest average at 3.19, but also demonstrate the largest amount of improvement of .52. The most noticeable difference was shown in the "Style" category, which showed more improvement in Spring 2012 than in Fall 2011 (.31 average points of improvement compared to .15). Additionally, the ratings for "Surface" category showed less improvement in Spring 2012 than in Fall 2011 (.15 points of improvement compared to .24). Finally, the results show that students improved more in the "Organization" category in Spring 2012 than in Fall 2011 (.40 compared to .25).

As Table 5, below, shows, the Spring 2012 data shows much greater statistical significance than the Fall 2011 data.

⁷ Note that all column headers and numerical values shown in Table 4 correspond to the same column headers and numerical values in Table 2.

Table 5*Statistical Analysis, Spring 2012*

Quantity	Pre Mean	Post Mean	p-value (two sided)
Holistic	4.210	4.7903	0.028
Thesis	2.597	2.9678	0.032
Org.	2.613	3.0161	0.004
Dev.	2.661	2.9677	0.030
Style	2.742	3.0484	0.034
Surface	2.871	3.0161	0.234
Presentation	2.677	3.1935	0.002
Mean (excluding Holistic) (n=62)	2.694	3.035	0.002

Table 4 shows statistically significant improvements at the .05 target level in all categories except for the “Surface” category (which also fails the less stringent significance test at .10). That is, because the p-values of the “Holistic,” “Thesis,” “Organization,” “Development,” “Style,” “Presentation,” and “Mean” categories fell below .05, we might say that improvements shown from pre- to post-intervention were in all likelihood not due to chance.

Interpreting two semesters’ worth of data. We saw dramatic differences between fall and spring data. The Fall 2011 data shows statistically significant improvements only in terms of “Development,” while the Spring 2012 data demonstrates significance in all categories other than “Surface.” If we work solely from the results of our ratings prior to the two-sided test—again, this is the test we use when we can’t say with certainty if our results will show improvements or deterioration—we might say that that writing center tutoring during 2011–2012 may have contributed to improvement in writing in terms of both “Holistic” and several specific category measures. If we factor in p-values, however, we find much higher rates of statistical significance in Spring 2012 than in Fall 2011. As noted above, these findings would seem to indicate any improvements in student writing are not due to chance but, instead, writing center intervention. However, because we did not collect con-

trol data in these studies, we cannot make any such claim without some qualification.

A number of faculty members from fields in which quantitative studies are common commented that the data showed it was *possible* but did not *establish* that writing center intervention was responsible for the improvement demonstrated in the study. Two alternate hypotheses explaining the improvements were proposed. The first alternate hypothesis explains the changes from pre- to post-intervention averages as a result of the expected maturation and learning process for first-year students. All of the papers had been gathered from freshman-level courses. We expect students to enter college at a certain level and then improve during the course of a year. With no control data, we could not claim that students who came for a writing center appointment improved more between draft and final version than students who didn't have a tutoring appointment in our center. Perhaps all students can be expected to make such improvements. Some data from Fall 2011 and Spring 2012 do, in fact, seem to support this alternate hypothesis. The post-intervention average on the nine-point holistic scores for Fall 2011 was 4.29; the pre-intervention average on the same scale in Spring 2012 was very similar at 4.21. Could it be that the educational environment in general had produced a first-year class that could score, on average, 4.29 at the end of the Fall 2011 semester and that those students remained at that level when we gathered pre-intervention papers from them in Spring 2012? Were we seeing evidence of the positive effect of the entire educational environment—including the specific writing courses the students were enrolled in—rather than any specific effect produced by writing center intervention? These are confounding variables indeed!

The second alternate hypothesis held that the improvements were the result primarily of self-selection bias. That is, students who cared enough about their writing to devote time and attention to improving their drafts were also more likely to seek tutoring, a phenomenon that would impact our study results. This explanation suggests that these students' scores improved because they were intrinsically motivated—that is, they visited the writing center of their own volition. Intrinsic motivation has been shown to be key to writing performance, as per Heather Robinson (2009). Robinson shows that after three visits, students visiting the writing center demonstrate increased intrinsic motivation and stronger writing performance. While we did not count the number of student visits, the methodology of our 2012 study—which relied on students not only coming to the writing center voluntarily, but also voluntarily bringing in a revised copy of the paper—would seem

to lend some credence to our second alternate hypothesis, and, perhaps, Robinson's work.

Intervention vs. Non-Intervention Study I

In an attempt to test for these two alternate hypotheses, we devised a plan to add control data to the study at the end of the Spring 2012 semester. Each semester, the English Department conducts its own internal assessment. The department's assessment requires all ENGL 101 teachers to contribute three randomly-chosen papers to a department-wide assessment that rates each paper according to three of the department's Student Learning Outcomes (SLOs) for ENGL 101. Each teacher contributes a final draft of a research paper usually written toward the end of the semester. Randomness in the selection of students is achieved by a simple method: Each teacher is asked to look at their class roll and select papers from students #3, 9, and 17. If student #9 doesn't turn in a paper, the teacher is to select the paper from student #10, and so on. The three SLOs addressed by the assessment were as follows:

- SLO 1: Ability to comprehend and analyze language
- SLO 2: Ability to express oneself clearly and effectively
- SLO 3: Ability to comprehend, analyze, and critically evaluate information

Each paper receives a rating between 1 and 3 points on each of the above areas. A rating of 2 is given to papers considered to be "on target" for that area, while a rating of 1 is "below target" and a rating of 3 is "above target."

In Spring 2012, we asked the English Department to add one new piece of data on the intake sheet for each paper. A check mark in a box accompanying the question "Was this paper seen at the writing center?" would put the paper in the intervention group. Papers without this check mark would go in the non-intervention group. When all papers were gathered, there were 103 papers in the non-intervention group and 48 in the intervention group. If the positive results in our pre-intervention vs. post-intervention studies were the result of the writing center intervention and not explainable primarily by either of the two alternative hypotheses, we would expect the intervention group to outperform the non-intervention group, and thus the results of this new study would corroborate the results of our two-semester writing center study.

However, when all of the data were entered and averaged, there was no discernible difference between the ratings for the two groups. In fact, the non-intervention group out-performed the intervention group

slightly overall, with the non-intervention group averaging 1.89 on all three measures and the intervention group averaging 1.88 on all three measures. The cumulative average and the average for each SLO are shown in Table 6, below:

Table 6

Assessment Data, Intervention vs. Non-intervention Groups, Spring 2012

Group	SLO1	SLO2	SLO 3	Mean
Non-intervention (n=103)	1.81	2.03	1.76	1.89
Intervention (n=48)	1.86	2.00	1.72	1.88

On SLOs 2 and 3, the non-intervention group out-performed the intervention group by small amounts. The intervention group demonstrated a higher average on SLO1, but any differences between the averages for the two groups are too small to refute the conclusion that the two groups performed very similarly. There was no statistically significant difference between the scores for the two groups, as shown below in Table 7:

Table 7

Statistical Analysis, Intervention vs. Non-Intervention Groups, Spring 2012

SLO	Non intervention Mean	Intervention Mean	p-value (two sided)
1	1.809	1.861	0.155
2	2.029	2.000	0.472
3	1.757	1.715	0.309

There was no significant difference between the intervention and the non-intervention test at the 5% level of significance. As we have seen, above, any results that do not meet this level of significance can be attributed to chance. As with the 2011–2012 studies, we ran two-sided

tests because we did not make the assumption that the intervention group was supposed to be better than the non-intervention group. (If we had run a one-sided test, the p-values would have been cut in half, but even then, all of the results would have failed to meet the 5% level of significance). We could not use the Spring 2012 data, then, to posit either a causative or correlative relationship between writing center intervention and improved writing.

These results motivated us to further refine our assessments. However, new data would not be available until we conducted a follow-up study in Fall 2012. In the meantime, we looked for reasonable hypotheses that would explain the control/test data we had gathered. Our science colleagues asked if the English Department’s SLOs were substantially different from the rating areas we had used in our writing center study, but a quick look at the SLOs shows too much overlap between the specific rating areas in the two studies to support this hypothesis. SLO2, for example, was broken into five sub-areas, each of which was rated according to the 1–3 point scale. Each of those criteria was closely aligned with at least one of our specific rating categories in the writing center study. Table 8, below, shows a correspondence between these sub-areas in SLO2 and five of the six categories in the writing center study:

Table 8

Corresponding SLO2 and Writing Center Assessment Categories

English Dept. Study Category	Writing Center Study Category
Establish a main point, focus, or argument	Thesis
Support the main point with reasons or evidence	Development
Organize and structure the project logically	Organization
Employ varied sentence structure, effective diction, and engaging style	Style
Conform to conventional mechanics, spelling, grammar	Surface

Some colleagues also asked if the raters in the English Department's intervention/non-intervention study were different or had been trained differently than the raters in the pre-/post-intervention studies, but neither was the case. Four of the six raters from the English Department study had been on the panel of readers for the pre-/post-intervention study, including the table leader.

An alternate hypothesis generated by the new data argues directly against the effectiveness of writing center tutoring. Students in the English Department's assessment scored almost exactly the same, whether they visited the writing center or not. Therefore, we might claim that tutoring has no positive effect on students' writing. Indeed, such a claim might be supported, in part, by our statistical analysis from Fall 2011, in which we saw statistically significant improvement in only two categories ("Development" and "Mean"). Two alternate hypotheses to refute these interpretations are equally possible, though. The first views the similarity in the non-intervention group and intervention group data as a result of self-selection bias. It could be that the lower-performing students had self-selected to visit the writing center, and their work improved as a result. As Williams & Takaku (2011) show, ESL students with low English writing proficiency "received higher grades in composition" than students who did not receive writing center tutoring, "regardless of their ESL or native-English speaker status" (p. 13). What was true for Williams & Takaku's ESL cohort may also be true of our intervention group, whose ratings may have been significantly lower than those of the non-intervention group had the intervention group not visited the writing center.

The second alternative hypothesis explains the similarity in the two groups' data as a result of a lack of sensitivity in the English Department's scoring system. It could be that most students' scores tended to cluster in the two-point range in the Department's three-point scoring system. A score of two merely represented a rating of "Meets Expectations." It could be that many of the papers in the test group improved, but not enough to move up to the standard for "Exceeds Expectations." If the rating scale had had at least five levels of sensitivity, it is possible that a number of scores for the test group would have improved because the scale would have offered more rating levels. While we cannot confirm either of the above alternative hypotheses, they are entirely reasonable interpretations of the data.

Writing Center Assessment: Year Two

During Fall 2012, we attempted to add control data to the pre-/post-intervention model we had used in 2011–2012. The new plan called for three specific instructors of English 101 to be involved in gathering papers rather than allowing students to self-select for the study. Each instructor would have both an intervention section and a non-intervention section. Students in the intervention sections would be required to visit the writing center after completing a draft of a research paper. Students in the non-intervention sections would not be required to visit the writing center, and if they chose to visit the writing center, their papers would be excluded from the study. Draft and final versions of the students' papers in both groups were collected. In each class section, the draft was collected approximately one week before the final paper was due. For ethical reasons, we did not ask the instructors to tell their non-intervention sections of English 101 that they were not allowed to visit the writing center.

When all papers were collected, there were 35 draft/final pairs in the non-intervention group and 50 in the intervention group. These papers were then read blindly by a panel of six faculty raters according to the same protocols used in Fall 2011 and Spring 2012. The results below may be counterintuitive, but there are reasonable hypotheses to explain them. On the nine-point holistic score, the 50 papers in the test group consistently out-performed the 35 papers in the control group. However, the intervention group's holistic score average rose less from draft to final than the non-intervention group's scores. The results for the holistic score are shown in Table 9, below:

Table 9

Holistic Ratings, Intervention vs. Non-Intervention Groups, Fall 2012

Group	Draft	Final	Differential
Intervention (N=50)	4.53	4.73	+0.20
Non-intervention (N=35)	3.32	4.11	+0.79

As Table 9 shows, the intervention group scored more than .6 points higher on their final papers than the non-intervention group. However, the intervention group's average holistic score improved only .2 points from draft to final, while the non-intervention group's average holistic

score improved by nearly .8 points. The most notable difference between the two groups may be the gap between their draft averages, however. The intervention group scored more than 1.2 points higher on their drafts than the non-intervention group. In fact, the intervention group's draft scores were more than .4 points higher than the non-intervention group's final scores.

Scores for the six categories followed the same pattern as the holistic scores. Table 10, below, shows that in every category, the intervention group scored higher than the non-intervention group. Additionally, the intervention group's draft scores were slightly higher than the final scores for the non-intervention group in all but the "Presentation" category.

Table 10

Category Ratings, Intervention vs. Non-Intervention Groups, Fall 2012

Category	Non-in- tervention Draft	Non-in- tervention Final	Interven- tion Draft	Interven- tion Final
Thesis	2.20	2.55	2.85	2.91
Organization	2.35	2.67	2.86	3.04
Development	2.27	2.79	2.86	2.95
Style	2.27	2.76	2.84	3.37
Surface	2.72	3.02	3.12	3.18
Presentation	2.41	3.11	2.73	3.19

These results demonstrate that the intervention group consistently out-performed the non-intervention group on both the holistic and category scores. However, because these results also demonstrate a higher level of improvement from draft to final version for the non-intervention group than for the intervention group, they seem at first to be at odds with the results from the previous year's study, which showed more improvement in holistic scores after visiting the writing center (.55 points in Fall 2011 and .58 points in Spring 2012 as compared to .2 points in Fall 2012).

Statistical analysis of this data (shown in Table 11, below) does little to help us explain these results:

Table 11

Statistical Analysis, Intervention vs. Non-Intervention Groups, Fall 2012

Category	Non-Intervention Draft-Final Change Mean	Intervention Draft-Final Change Mean	p-value (two sided)
Holistic	0.790	0.220	0.046
Thesis	0.353	0.040	0.125
Organization	0.324	0.180	0.235
Development	0.515	0.09	0.014
Style	0.485	0.527	0.752
Surface	0.309	0.060	0.119
Presentation	0.706	0.500	0.236
Mean (excluding Holistic)	0.449	0.233	0.062

As the above table shows, we compared the mean, or average, change from draft version to final version in all categories, including the holistic category, for both the intervention and non-intervention groups. In general, the mean improvement in the intervention group was lower than the mean improvement in the non-intervention group; the exception occurred when comparing the “Style” category, in which the intervention group improved slightly more than the non-intervention group. However, the differences between the two groups’ levels of improvement only met the 5% level of significance in two areas. At .0139, the p-value in the “Development” category indicates strong statistical significance, but at .0464, the p-value in the “Holistic” category is only slightly better than the .05 level needed to declare statistical significance. This indicates that any changes in the “Development” and “Holistic” categories were likely not due to chance. For all of the other areas, we found no statistically significant differences between the mean changes for the two groups, which means, again, that those changes were in all likelihood due to chance. We could conclude that the data from the Fall 2012 study does not, in general, demonstrate a meaningful difference

between the levels of improvement for the two groups. The exception, again, appears to be the category of "Development," which shows statistically significant improvement.⁸

Per the calculations of mean averages shown in Table 10, it is clear that the intervention group consistently out-performed the non-intervention group in all areas. However, per Table 11, the Fall 2012 results (with the exception of "Development") show results above the 5% significance threshold. These results raise the question of why raw data that appears to support the value of writing center intervention simultaneously fails the two-sided t-test. Why, in other words, do the means, or averages, we show in Table 10, above, seem to indicate that writing center intervention has a positive impact on writing, while the statistics we report in Table 11, above, seem to indicate that most improvements (with the exception of "Development") are probably due to chance? We are forced to conclude that the majority of the Fall 2012 results shown in Table 11 do not conclusively refute the null hypotheses—which is to say, we cannot discount other hypotheses for improvements in student writing.

One such hypothesis might be called the "dental hygienist effect." In the same way that many people put extra effort into brushing and flossing in the days leading up to a visit to the dental hygienist, it is possible that the students who were required to visit the writing center worked harder on their drafts than students who weren't required to visit. Thus, there could be a correlation between anticipation of a writing center visit and improvement on a draft. Another hypothesis might be called the "willing partner" effect. It could be that the instructors either consciously or unconsciously influenced the students in their intervention sections to perform better on the writing assignment than their non-intervention sections. This hypothesis is plausible because both the instructors and the students knew that they were involved in a study of the writing center and knew which group each class section was a part of.

Conclusions and Recommendations for Further Study

Writing center administrators want to see that quantitative assessments produce more than just interesting sets of data. They want to "close the

8 Perhaps we can posit, however tentatively, that "Development" is the trait most readily and most visibly impacted by writing center tutoring. This makes sense, given that our tutor training protocols privilege higher-order concerns such as development.

loop,” meaning that assessments need to be used to impact positively the work done in the center—even if those results, like ours, were mixed. At Coastal Carolina University, we have accomplished this goal in several important ways, most immediately through tutor training. Tutors are required to attend eight training sessions per year on various topics, and the results of our studies have augmented several recent training sessions. When the Fall 2011 study showed that students who visited the center showed the least amount of improvement in “Style,” we targeted this category in Spring 2012 training sessions. Average student improvement in this category subsequently improved in Spring 2012. Additional training sessions have focused on various issues in response to what we perceived as weak scores in the “Presentation” and “Organization” categories.

Assessment results have also helped us create more robust resources and renew our investment in extant ones. We have developed and uploaded to our web page a series of narrated PowerPoint lessons on APA, MLA, and Chicago formatting. Instructors and tutors use these resources to conduct in-classroom workshops with students; tutors use them to supplement tutoring sessions; and students access them on their own. Prior to these assessments, the English Department’s Composition Committee had considered dropping the University’s in-house first-year writing handbook. With assessment results in hand, however, that committee recently revised and expanded the handbook to include a chapter on Chicago formatting. At this point, we have not yet begun a rigorous study of the effects of these new resources, but we have some encouraging preliminary data. The new PowerPoints are being accessed hundreds of times each semester, and our *First-Year Composition Guide*⁹ is not only required in all first-year writing courses but is by far the most frequently used resource in the Writing Center. Tutors rely on it even more heavily than the *MLA Handbook* or the Purdue OWL when they provide assistance to students writing in MLA format.

Our assessments have also had at least one other important impact. Throughout the nearly four-year process of designing, conducting, and documenting our results, tutors have received periodic updates on this study. While Coastal Carolina University’s largest programs are in science and business fields, we draw the majority of our tutors from the English and other humanities field. In the process of learning about the study, these tutors have become more familiar with techniques, terms, and ways of thinking that they might not otherwise have been exposed

⁹ The title of this in-house guide was changed to *Coastal Writers’ Reference* for the third edition, which was published in Fall 2014.

to. Our tutors now know the difference between the term *significance* as it is typically used in the humanities and that same term as it is applied to statistics. These tutors have become more familiar and comfortable with concepts and terms such as *intervention*, *control group*, and *hypothesis*, and they learned from the development of a research report that is similar to the kinds of writing they see from our campus's many science majors.

Our next goal is to design a study to help us interpret and build from the data we have already gathered. As we anticipate this follow-up research, we confirm our support of what Cindy Johaneke advocates: a closer attention to context and a richer mix of strategy. In *Composing Research: A Contextualist Paradigm for Research and Composition* (2000), Johaneke argues that researchers should select methodologies to fit specific research questions and should not limit themselves to any specific methodology because "[i]n a Contextualist Research Paradigm, one kind of research is not more valuable than another, and one kind of evidence does not guide our quests" (p. 207). Like Johaneke, we are not primarily interested in quantitative work for its own sake. Rather, we seek to situate quantitative studies within the richer context of knowledge-building. To this end, we plan to augment future studies with pre- and post-intervention interviews and focus group meetings with all available stakeholders: students, tutors, and faculty raters. This triangulation will help us tell the story that our data support. It might also help us better understand phenomena we are unable to control for—the "dental hygienist" effect, for example—and that which we see, but do not understand. Why, for example, does the "Development" category demonstrate statistical significance across each study? Is development privileged in tutor training? Is it easiest to tutor? Or is it most obvious to raters? Further, and perhaps more importantly, how can we augment tutor training to bring about stronger results in the other categories studied?

The numbers from the Fall 2011 and Spring 2012 studies suggest, at points, positive correlations between writing center intervention and improved student writing. Unpacking those first two data sets even further, it is important to note that the numbers from the Fall 2011 and Spring 2012 studies loosely align with the numbers from Niiler's 2005 study. The improvements demonstrated from pre- to post-intervention drafts in these three studies range from .55–.58 points on five-point scale. These results suggest, albeit tentatively, that the pre-/post-intervention model and intervention/non-intervention models could be valid tools for measurement across institutions. Within the context of this study, we have seen that our work is data-supported and aggregable. It is also replicable, with protocols that can be adopted by and adapted

to the varying needs of myriad institutions. Unsatisfied with counting clients and reporting the results of client satisfaction surveys, we created and implemented two pre-/post-intervention studies, and two intervention/non-intervention studies. We identified assessable outcomes, per Macauley (57); gathered writing samples; trained raters; gathered data; performed relevant statistical analyses; and found ourselves by turns pleased and confounded by the results. Likewise, if as a discipline we can construct, share, and enact a set of “best practices” for outcomes assessment, we may be able to ensure the institutional health and viability of writing centers and create compelling justifications for new ones.

We might also consider the primary audiences for future outcomes assessments. That audience is not limited to administrators and accreditation committees, but includes the entire writing center community. Assessments of writing center outcomes speak most clearly to us, in our own spaces, as we work to reflect on what we’ve done with an eye to doing it better. Understood as reflective practice, outcomes assessments help us articulate our own best practices and train our own tutors to do their best work. Further, the work of assessing outcomes can help directors articulate more robust outcomes—and, by extension, daily practice. There are important differences, for example, between stating that a writing center exists to “help students develop the writing skills necessary for academic success” and saying that it will help students improve in terms of thesis, organization, style, and surface. The former statement is a vague wish; the latter are actionable, measurable categories.

A final note: As with Niiler’s prior studies, the current project both lends credence to and makes problematic the claim that writing center intervention helps students’ writing improve. This kind of ambiguity should be expected and even welcomed as more writing center researchers are encouraged to embrace RAD methodologies. Even with the most rigorous of experimental protocols, we must learn to live with some uncertainty. In his discussion of “experimental method” in *The Making of Knowledge in Composition* (1984), North says:

Experimental knowledge, no matter how carefully or rigorously tested, remains relative, a probability.... No number of proofs that some apparent relationship between variables is not the result of chance adds up to the final proof that it is, in fact, the result of any particularly posited connection. (pp. 151–152).

While we are aware of these theoretical limits, we encourage other writing center professionals to conduct their own data-driven research. We hope our work has shown that there is much meaning to make, in terms of both our local and disciplinary practices.

References

- Babcock, R. D., & Thonus, T. (2012). *Researching the writing center: Towards an evidence-based practice*. New York, NY: Lang.
- Barnett, R. W., & Blumner, J. S. (2001). *The Allyn & Bacon guide to writing center theory and practice*. Boston, MA: Allyn and Bacon.
- Bell, D. C., & Frost, A. (2012). Critical inquiry and writing centers: A methodology of assessment. *The Learning Assistance Review*, 17(1), 15–26. Retrieved from <http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=EJ973400>
- Bredtmann, J., Crede, C. J., & Otten, S. (2013). Methods for evaluating educational programs: Does writing center participation affect student achievement? *Evaluation and Program Planning*, 36, 115–123. Retrieved from doi: 10.1016/j.evalprogplan.2012.09.003
- Coastal Carolina University Composition Committee (2013). *Coastal Carolina University first-year composition guide* (3rd ed.). Southlake, TX: Fountainhead Press.
- Coastal Carolina University Writing Center (2015). Online tutorials. *Coastal.edu*. Retrieved from <https://www.coastal.edu/writingcenter/tutorials.html>
- Driscoll, D., & Perdue, S. W. Theory, lore, and more: An analysis of RAD research in *The Writing Center Journal*, 1980–2009. *Writing Center Journal*, 32(1), 11–39.
- Griswold, G. (2003). Writing centers: The student retention connection. *Academic Exchange Quarterly*, 7(4), 277–81.
- Haswell, R. H. (2005). NCTE/CCCC's recent war on scholarship. *Written Communication*, 22(2), 198–223.
- Henson, R., & Stephenson, S. (2009). Writing center consultations can effect quantifiable change: One institution's assessment. *Writing Lab Newsletter*, 33(9), 1–5.

- Johanek, C. (2000). *Composing research: A contextualist research paradigm for rhetoric and composition*. Logan: Utah State University Press.
- Irvin, L. L. (2014). What a difference three tutoring sessions make: Early reports of efficacy from a young writing center. *Writing Lab Newsletter*, 39(1–2), 1–5.
- Jones, C. (2001). The relationship between writing centers and improvement in writing ability: An assessment of the literature. *Education*, 122(1), 2–20.
- Kail, H. (1979). Evaluating our own peer tutoring programs: A few leading questions. *Writing Lab Newsletter*, 7(10), 2–4.
- Lerner, N. (1997). Counting beans and making beans count. *Writing Lab Newsletter*, 22(1), 1–4.
- Lerner, N. (2012). Of numbers and stories: Quantitative and qualitative assessment research in the writing center. In E. Schendel & W. J. Macauley (Eds.), *Building writing center assessments that matter* (pp. 106–113). Boulder: University Press of Colorado.
- Macauley, W. J. (2012). Getting from values to assessable outcomes. In E. Schendel & W. J. Macauley (Eds.), *Building writing center assessments that matter* (pp. 25–56). Boulder: University Press of Colorado.
- McCracken, N. (1979). Evaluation/accountability for the writing lab. *Writing Lab Newsletter*, 3(6), 1–2.
- Mohr, E. (1995, 27–20 Sept.). Researching the effectiveness of a writing center. Presentation at International Writing Centers Conference, St. Louis, MO. Retrieved from <http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED419233>
- Newmann, S. (1999). Demonstrating effectiveness. *Writing Lab Newsletter*, 23(8), 8–9.
- Niiler, L. (2003). The numbers speak: A pre-test of writing center outcomes using statistical analysis. *Writing Lab Newsletter*, 27(7), 6–9.

- Niiler, L. (2005). The numbers speak again: A continued statistical analysis of writing center outcomes. *Writing Lab Newsletter*, 29(5), 13–15.
- North, S. M. (1984). Writing center research: Testing our assumptions. In G. Olson (Ed.), *Writing centers: Theory and administration* (pp. 24–35). Urbana, IL: NCTE.
- North, S. M. (1987). *The making of knowledge in composition: Portrait of an emerging field*. Upper Montclair, NJ: Boynton.
- Roberts, D. H. (1988). A study of writing center effectiveness. *Writing Center Journal*, 9(1), 53–60.
- Robinson, H. M. (2009). Writing center philosophy and the end of basic writing: Motivation at the site of remediation and discovery. *Journal of Basic Writing*, 28(2), 70–92.
- Schendel, E., & Macauley, W. J. (2012). *Building writing center assessments that matter*. Boulder: University Press of Colorado.
- Schmidt, K. M., & Alexander, J. E. (2012). The empirical development of an instrument to measure writerly self-efficacy in writing centers. *Journal of Writing Assessment*, 5(1), 1–10. Retrieved from <http://www.journalofwritingassessment.org/article.php?article=62>
- Thompson, I. (2006). Writing center assessment: Why and a little how. *Writing Center Journal*, 26(1), 33–61.
- Waldo, M. L. (1987). More than 'first-aid': A report on the effectiveness of writing center intervention in the writing process. *Issues in College Learning Centers*, 5, 12–22.
- Williams, J. D., & Takaku, S. Help-seeking, self-efficacy, and writing performance among college students. *Journal of Writing Research*, 3(1), 1–18.
- Yeats, R., Reddy, P., Wheeler, A., Senior, C., & Murray, J. (2010). What a difference a writing centre makes: A small-scale study. *Education and Training*, 52(6/7), 499–507.

About the Authors

Scott Pleasant has served as the Writing Center Coordinator at Coastal Carolina University since 2010. His primary research focus during that time has been on quantitative writing center assessment, but he has also published and presented on tutor training, writing center promotion, and other issues. In addition to his work with the Writing Center at CCU, he teaches research methods and other courses in the Interdisciplinary Studies program and coordinates the assessment activities for his college. With Joe Oestreich of the CCU English Department, he recently co-wrote *Lines of Scrimmage: A Story of Football, Race, and Redemption* (University Press of Mississippi, 2015).

Luke Niiler is an associate professor of English and director of the First-Year Writing program at the University of Alabama. He formerly directed UA's Writing Center. His teaching and research interests include composition and writing centers.

Keshav Jagannathan joined Coastal Carolina University in 2005. He is an Associate Professor of Statistics in the Department of Mathematics and Statistics. Dr. Jagannathan's research interests are varied, from statistical distribution theory to pedagogical advances in mathematics and statistics. He is currently responsible for the mathematics placement at Coastal Carolina University.