2019 SWCA Plenary Address
Producing Better Writing

Jo Mackiewicz

I started thinking about this presentation months ago when Scott[1] first asked me to appear as a speaker at this conference. At that time, I thought I had something to say about what I thought was a conspicuous lack of research on writers' revisions after writing center conferences. I thought I wanted to say that it is high time we acknowledge that our aim should be to improve writing as well as writers. I wanted to wield a metaphorical sword and confront Steven North's proclamation. That was at the beginning. And so the blurb for the talk, sent to Scott months ago, went as follows:

> In this plenary session, I discuss a topic I think receives little attention in writing center conversations: our role in helping student writers to improve their writing. I argue that improving writing is a worthwhile and measurable endeavor and that doing so offers opportunities for meaningful and valued assessment.

In the months that followed, I dug into the extant research on writing centers and revision and, indeed, revision after instruction in general. In the process, I changed my mind about the importance of focusing on writing over and over again. In this talk, I'd like to map out that cognitive journey for you. Perhaps you'll find that I ended up where you already are—and maybe you didn't need to run a cognitive marathon to get there. I think I have ended up in a better place than where I started. I have a

---

[1] Scott Pleasant, chair of the 2019 Southeast Writing Centers Association conference.

better understanding of the return on investment of investigating revision.

But first, let me backtrack a bit. I need to describe what I saw (and still see) as a big problem in writing center research: hardly any research examines the differences in pre-conference and post-conference papers. I think this situation is a problem because it means we have little evidence that the writing center work we do manifests in writers' deliverables. What, after all, does it mean to be a "good writer" if you can't demonstrate "good writing"? A pragmatic problem arises from this lack of research too: writing centers cannot show administrators that their work positively affects writing in the short term, let alone in the long term. Writing centers miss opportunities, then, to underscore the important work they do, the one-to-one conversations aimed at helping writers meet their goals.

Before I describe the some of the few studies on revision that exist, it's a good idea, I think, to acknowledge two main reasons that writing center scholars have not studied revision very often:

1. Practicality: It is difficult to operationalize types of revision. It can be time-consuming to code and compare changes from draft to revision. It is difficult to code conference talk and map that conference talk to revisions, particularly over time.
2. Philosophy: (i.e., "We make better writers"). In other words, because writers are our concern, our research need not focus on their writing.

When I began developing this talk months ago, I underestimated the practical challenges of conducting research on revision. I had done a fair amount of research involving coding, and I'm generally of the mindset that it is possible to operationalize most any discourse-related construct, revise the coding scheme for inter-rater reliability, and thus determine the frequency and location of various linguistic items. But tracking revision, even from one draft to another, is challenging.

In addition, when I started this journey, I think I failed to recognize fully that, for writing centers at least, improving writing is important to the extent that the improvement demonstrates improvements in writers. That is, I was before more of the mind that writing improvement in and of

itself has value. I think for the most part I'm of the mind that studying revision is important because it is a *direct* gauge of whether people have improved as writers. And, it also provides opportunities to show the value of the support that writing centers provide.

In this talk, I overview some critical studies—ones that have attempted to do the hard work of operationalizing writing revision. Some of these studies have tried to tie writing center discourse to revisions in order to show that writing center conferences had an effect on writers' work, as opposed to other potential inputs such as instructors' feedback. As I move through these studies that have shaped my current thinking, I also discuss some limitations of those studies. Then, I posit one potential method of studying changes from draft to revision.

Critical Studies on Revision

This is the point where we start to get in the weeds.

Faigley and Witte, "Analyzing Revision"

Lester Faigley and Stephen Witte's (1981) study of revision in 6 inexperienced student writers, 6 advanced student writers, and 6 expert adult writers was groundbreaking. The study took place over three days. On the first day, writers were given the topic; on the second day, they wrote their essays; on the third day, they revised. (Sounds a little biblical, doesn't it?)

 Faigley and Witte found this:

- Not surprisingly, inexperienced writers made the most surface changes.
- 24% of advanced students' changes were meaning changes (what Faigley and Witte call "macrostructure" and "microstructure").
- 34% of expert adults' changes were meaning changes (i.e., macrostructure or microstructure).

In the second part of their study, they asked expert adults to revise three inexperienced students' essays as if they were their own. They compared the changes the experts made to the changes that the students had made.

They found that 65% of the changes that the expert writers made were macrostructure changes.

The findings of their first study are particularly helpful to the study of revision because they indicate the depth of revision that writers of different experience will carry out given the same (albeit unnatural) writing condition. However, the main contribution of this study is its coding scheme for revision. Other researchers who have attempted to code and quantify revision have used it. However, as I'll discuss more, using this coding scheme isn't as simple as some researchers have made out.

Stay, "When Re-Writing Succeeds: An Analysis of Student Revisions"

Using Faigley and Witte's scheme, Byron Stay studied 20 students' writing. These students had to visit the writing center after earning a D or D+ in freshman comp. The students had to write two papers for evaluation by writing center instructors. They had to write in the writing center and save all the drafts for their papers.

Stay's big contribution to revision research was that he evaluated each revision (at the sentence level it seems) as positive, negative, or neutral. (Note, however, that he gave no measure of inter-rater reliability for this evaluation.) He reported findings in terms of words changed per 100 words.

Here's what Stay found:

His findings differed from those of Faigley and Witte. Stay found that students made a *large* number of macrostructure changes (when compared to Faigley and Witte's study). He writes, "One reason for the high number of macrostructure changes is that many students voluntarily revised their first draft extensively after consulting with *an instructor*… [italics added]. The relatively low number of surface changes may be partially attributed to the instructor who helped keep the students' attention fixed on larger questions of organization, especially in early drafts" (p. 23). However, of all revisions, just 57% of the changes were evaluated as positive: 10.5% were negative and 32.5% were neutral.

Stay's study is one that gets cited quite a bit, along with Faigley and Witte's study. However, the participants' situation (writing to pass a course, writing in the writing center, receiving feedback from FYC instructors that was not accounted for in the study) puts into question the results for typical writing center conferences—along with other methodological problems.

Bell, "Better Writers: Writing Center Tutoring and the Revision of Rough Drafts"

Bell used Faigley and Witte's scheme also. James H. Bell's big contribution was that he classified each conference according to Thomas Reigstad's typology: tutor-centered, structured participation, collaborative, and student-centered. Doing this, he could match conference type to extent and quality of revision. He performed two studies.

Study 1: The first employed four experienced tutors working with four student writers. The students' draft and final papers were compared and evaluated in terms of Faigley and Witte's scheme. The quality of each change was evaluated as positive, negative, or neutral. Note that Bell used percentage of agreement for inter-rater reliability (not Cohen's kappa); the inter-rater percentage was 88% for both the textual changes and the quality assessments.

In study 1, Bell found:

Of the changes students made, 69% were in line with students' objectives, and 79% were improvements, but he says, "there is no clear indication whether students became better writers, for nearly all changes to texts were made during sessions" (12). In other words, the changes were those directed or suggested with real-time guidance. He also found that the conferences were tutor-centered and focused primarily on surface changes.

In study 2, Bell studied a professional tutor working with 11 undergraduate writers. Tutoring sessions were classified as assignment-assistance (i.e., focused on completing the assignment) or instructional (i.e., focused on teaching something to make the student a better writer). The professional tutor conducted instructional conferences.

In study 2, Bell found:

1. 32% of in-line changes (i.e., surface changes) were made during the sessions; 68% were made after by the students.
2. 86% of the changes made during the sessions were in line with the stated objectives of the conference (although it's not exactly clear how Bell determined what the objective was).
3. 65% of the changes were in line with what the tutor had tried to teach in the session.
4. 98% of the changes made during the conference were positive, and 82% of the changes made after the conference were positive.

The conclusion: a trained, experienced tutor is better able to provide feedback so that students can implement it after the conference; they can negotiate and maintain an agenda; they better help writers make positive changes. In relation to Faigley and Witte's scheme, Bell found this: The professional tutor's suggestions were mainly macrostructure additions. The peer tutors initiated almost no macrostructure changes.

Williams, "Tutoring and Revision: Second Language Writers in the Writing Center"

Jessica Williams studied 4 L1 English tutors and 5 L2 writers. One of Williams's big contributions was that she linked "episode problematicity" (i.e., tutor suggestions, directives, notations of a problem/error, writer requests for assistance or notations of problems) to draft-to-draft changes. Another contribution was that she measured revision in a more straightforward way: (1) T-unit (i.e., minimal terminable unit) and (2) greater than T-unit. As a reminder, a T-unit is a main clause plus its dependent clauses (Hunt 20). For example,

- There was a man next door, and he was a police officer. = 2 T-units
- There was a man next door who was a police officer. = 1 T-unit

She found that writers made many changes that were "not attributable to anything that went on during the session" (181). The changes might have stemmed from other input: feedback from teachers, peers, or from the writers themselves as they generated more ideas.

Williams lists six main findings:

1. The focus of the discussion is usually the focus of the revision.
2. Surface-level features discussed during the session are more likely to get revised than text-based (i.e., meaning) problems.
3. Issues explicitly addressed by the tutor are more likely to be revised than those that receive more implicit treatment.
4. Writer response to tutor suggestions and explanations is predictive of the impact of that advice on revision. That is, if the writer noted the suggestion or explanation, the related change was likely to appear in the revision; if the writer resisted the suggestion/explanaItion, the related change was not likely to appear in the revision. If the writer offered a minimal response (e.g., "mmhm") or nonverbal backchannel, the related change was unlikely to appear in the revision (186).
5. Text-based (i.e., meaning) revisions that can be traced to writing center discussion are associated with some interactional features of writing center conference talk—features that tutors have some control over:
   a. Negotiation episodes (especially extended)
   b. Active writer participation in negotiation episodes
   c. Tutor's clarification of critical features (e.g., tutor helps writer clarify what he or she is trying to say)
   d. Tutor's sustained emphasis on goals
   e. Tutor's organization of the task
   f. Tutor's modeling of writing and revision strategies (e.g., outlining).
6. Revision does not always lead to higher-rated essays. (Raters used A–F holistic scale.)

Williams concludes: "It is likely that whatever their approach, tutors cannot really foster better writers directly. Rather, writers become better writers by working on their texts" (196). It's a subtle point that Williams is making here, I think. I believe what she is pointing out is this: Tutors' engagement gets writers to engage, and through that engagement, writers improve in their writing ability.

Van Horne, "An Activity-Theory Analysis of How College Students Revise after Writing Center Conferences"

Samuel Van Horne's dissertation study uses activity theory and Faigley and Witte's scheme to examine 11 students' revisions after 10 conferences. The big contribution from Van Horne's study was his method of analysis. He made some improvements on earlier studies:

1. Gave student a voice-recorder with instructions to record any follow-up conferences about the paper.
2. Gave student a structured journal and asked that he/she fill it out after each session of revision.
3. Used Word's versions and tracked changes to keep track of revisions.
4. Observed the student revise in order to learn about the decision-making process.
5. Interviewed the tutor and the student's instructor (about the writing center).
6. Differentiated between direct and indirect suggestions.

Two groups of students used the writing center: those who had specific goals for their conferences (i.e., those who had specific assignments) and those who did not.

Van Horne's findings included these:

1. Students without a specific agenda did not act on suggestions about how to expand on their ideas.
2. Like Williams and contrary to Stay, Van Horne found students were more likely to address surface errors than macrostructure or microstructure changes. Sometimes they fixed these surface errors but did not address meaning-related problems.
3. But Van Horne notes that tutors rarely made macrostructure-related suggestions.
4. Also, like Williams, Van Horne found that students revised the aspects of writing that they discussed with their tutors.
5. Finally, students incorporated indirect suggestions *only* when those changes were in line with what they believed their instructors wanted.

Problems with Faigley and Witte's Scheme and Subsequent Studies

Any careful analysis of the Faigley and Witte scheme reveals the problems that would come from trying to code revisions with it. Williams is the only writing center researcher who explicitly points to the elephant in the room: "Although authors report success with this system, few researchers in L2 writing have adopted the full array of categories. This may in part be because of the difficulty in reaching acceptable interrater reliability" (175). Williams goes on to ask, "What is the best way to express how and how much a text has been revised? How do we differentiate among the effects of revision? How can we measure the extent to which revision has resulted in improvement in the quality of the text?" (175). These questions are just as relevant today as they were in 2004, when Williams published her study.

Trying to answer these questions depends in part on trying to determine the shortcomings of earlier studies:

1. Faigley and Witte distinguish between surface changes and meaning changes (i.e., macrostructure and microstructure) changes. But it's not easy to say what's altered meaning and what isn't. Even small changes can alter meaning (e.g., changing "I have few friends" to "I have a few friends").

2. Although they don't state it explicitly and often use the term "unit," Faigley and Witte and other subsequent studies have analyzed changes at the sentence level. For example, in describing their method of counting revisions, Faigley and Witte say that one macrostructure change, an addition of seven sentences, would be counted seven times (405). One problem with this method, of course, is that sentences can vary greatly in length. One added sentence does not necessarily equal another.

3. Even though Faigley and Witte and other researchers who use their scheme seem to analyze revision in terms of sentences, they report results in terms of words changed per 100 or 1000 words. But it's not clear how the number of words changed were counted. For example, if a sentence is revised (say, through some permutation) and then moved to a different place in a paragraph, how are the words changed counted? Are words changed calculated from the place the words used to be and again in the spot they are placed? What about the words that have undergone

permutation? Are words changed measured per sentence (or another unit) or overall?

A Potential Method

There has to be a better way to study revision. The studies that I've reviewed have manifested characteristics of what I have in mind for a revision study.

1. Ask writers' instructors to provide feedback in written comments (Word's comments, for example). Ask students to use that version of the paper to work on their revision. Doing so tracks non-tutor advice.
2. Ask the tutor and writer to work on a computer during a face-to-face (in person or synchronous online) conference. Capture the audio and screen with software such as Camtasia.
3. Code tutoring strategies (Mackiewicz and Thompson) or other items of interest in the conference talk. Carol Severino and Shih-Ni Prim, in a longitudinal case study of one L2 writer's development, coded online tutor comments as (1) direct corrections; (2) noting an error exists; (3) explanations; (4) questions; (5) suggestions; (6) providing options.
4. Code tutoring strategies (or other item of interest) in the instructor's written feedback. It's important to differentiate between instructor and writing center input.
5. Ask writers to use screen-capture software to collect the changes they make from original to revision. Don't ask them to think aloud as they revise.
6. Ask writers to audio record any other discussions that they have about their paperss (instructor, peer).
7. Ask writer to review the screen-capture files immediately after revising and record a think-aloud as they watch themselves revise. Ask them to articulate why they made the changes they made. Writers can overlay their audio commentary on the screen-capture by screen-capturing the original screen-capture.[2]
8. Interview each writer and review the screen-capture + audio commentary. Ask further questions to elucidate their revision choices and motivations.

---

[2] At the conference, I played an example screen-capture video that I recorded as I drafted and revised this talk.

9. Use a data-analysis method like Williams's to gauge extent of revision: Williams describes her inter-rater tested method (91% agreement), which involved coding T-units into three types:
    (i) T-units that remained unchanged from the first to second draft, that is, the same text in the same sequence (i.e, no change)
    (ii) Those in which the elements of text were rearranged or slightly changed (i.e., slight change or small-scale change)
    (iii) Those in which larger chunks of text, at the level of the clause or larger, were added or changed" (i.e., substantial changes) (178).

   Or, use a data-analysis method like Severino and Prim's to measure complexity, accuracy, and fluency (CAF):

    (i) Fluency = total words per T-unit
    (ii) Accuracy = Errors per T-unit and error-free T-units
    (iii) Complexity = words per T-unit, words per clause, clauses per T-unit

10. Tie the writer's revisions to the conference's tutoring strategies and the instructor's feedback.
11. Rate each change in quality from original to revision, for example, successful, unsuccessful, deletion.

As I mentioned earlier, it would certainly be possible to code conferences for items other than tutoring strategies and then try to tie those items to subsequent revisions. For example, Williams tied students' requests to revisions as well. Bell analyzed direct versus indirect suggestions (i.e., telling and suggesting strategies versus hinting strategies). And as I mentioned, Severino and Prim used yet another scheme to code tutor feedback.

It would be possible to operationalize the domain of each directive or suggestion as well: surface-level, microstructure, etc. Actually, I think that Faigley and Witte's scheme might actually work better in terms of classifying the domain of tutors' feedback than it does for classifying writers' revisions.

It's Worth It

In addition to subjective and indirect measures such as conference-satisfaction ratings and writers' assessments of their progress, data from a study like the one I've sketched would show the support for writers that writing centers provide.

The quantitative data arising from such a study—even a study of one or two writers—would provide direct evidence of the kinds of short-term gains that writing centers enable. And, as Williams noted, "although one cannot directly extrapolate from short-term draft-to-draft change to long-term development, in the absence of demonstrated short-term revision, long-term improvement seems unlikely" (174). Once the researcher has developed the coding schemes and achieved viable inter-rater reliability, the coding could move along fairly quickly. The time invested would come from meetings with students about their screen-captures.

In addition, developments in writing research using keystroke logging and eye-tracking may facilitate revision research. My colleagues at Iowa State have used these technologies to study the extent to which students' assessment of their own writing processes positively impacts learning how to write in an L2 (Ranalli, Feng, and Chukharev-Hudilainen).

They found that using these technologies and talking about the data those technologies facilitated (1) positioning the two L1-Chinese students in their L2-writing development; (2) identifying and addressing problems related to writing process: planning, formulation, and revision; and (3) revealing motivational issues that hampered students' development.

Such technologies would be particularly useful, I think, for people who are interested in the cognitive process of revising, the moment-to-moment decision making and error correcting that go on as people write. These technologies are not necessary to study revision—they're just worth exploring. Perhaps they could ease the burden of data collection and analysis in longitudinal studies.

So there it is. I've ended up in a different place in my thinking than when I began. I started with the mindset that for whatever reason, writing center scholarship had eschewed empirical research on revision. Now, I have a better understanding. I know that such research is scarce, but that

it is becoming less so. As evidence, I point to Severino and Prim's longitudinal study again.

But I've also come away feeling like such research is quite feasible. More than that, I am more confident that the return on investment is worthwhile—both in terms about what we learn about the writing center's contribution to writers' gains and in terms of the writing center's ability to provide evidence of those contributions.

Thank you.

Works Cited

Bell, James H. "Better Writers: Writing Center Tutoring and the Revision of Rough Drafts." *Journal of College Reading and Learning*, vol. 33, no. 1, 2002, pp. 5–20.

Faigley, Lester, and Stephen Witte. "Analyzing revision." *College Composition and Communication*, vol. 32, no. 4, 1981, pp. 400–14.

Hunt, Kellogg W. *Grammatical Structures Written at Three Grade Levels*. No. 3. Champaign, IL: National Council of Teachers of English, 1965.

Mackiewicz, Jo, and Isabelle Thompson. *Talk about Writing: The Tutoring Strategies of Experienced Writing Center Tutors*. 2nd ed., Routledge, 2018.

Ranalli, Jim, Hui-Hsien Feng, and Evgeny Chukharev-Hudilainen. "Exploring the Potential of Process-Tracing Technologies to Support Assessment for Learning of L2 Writing." *Assessing Writing*, vol. 36, 2018, pp. 77–89.

Reigstad, Thomas J. "The Writing *Conference*: An Ethnographic Model for Discovering Patterns of Student-Teacher Interaction." *The Writing Center Journal,* vol. 2, no. 1, *1982,* pp. 9–20.

Severino, Carol, and Shih-Ni Prim. "Second Language Writing Development and the Role of Tutors: A Case Study of an Online Writing Center 'Frequent Flyer.'" *The Writing Center Journal,* vol. 35, no. 3, 2016, pp. 143–85.

Stay, Byron. "When Re-Writing Succeeds: An Analysis of Student Revisions." *The Writing Center Journal*, vol. 4, no. 1, 1983, pp. 15–28.

Van Horne, Samuel Alexander. *An Activity-Theory Analysis of How College Students Revise after Writing Center Conferences*. 2011. University of Iowa, PhD dissertation. ir.uiowa.edu/etd/1187/.

Williams, Jessica. "Tutoring and Revision: Second Language Writers in the Writing Center." *Journal of Second Language Writing*, vol. 13, no. 3, 2004, pp. 173–201.