

Tinkering with Comments: Tailoring Practice by Spying on Written Artifacts

Vicki R. Kennell, PhD, Josh Weirick, & Amy Elliot
IWCA, Chicago, Illinois, November 2017
[Notes to accompany workshop presentation slides.]

Slide 1: Vicki

Good afternoon, and welcome to our panel presentation. I'm Vicki Kennell, the ESL Specialist at the Purdue Writing Lab. With me today are Josh Weirick, a graduate student in the Linguistics PhD program and Amy Elliot, a graduate student in Literature and a tutor in the Writing Lab. We're going to talk to you today about how spying on the written artifacts left during asynchronous tutorial sessions and writing groups can help writing centers tailor their programs to create improved training opportunities for tutors and to better meet the needs of writers. You might be wondering "What about the soldiers?" Well, listen closely, and I promise they'll appear at some point . . .

Slide 2:

Our talk today has 3 main parts. I will start by talking about comment coding generally and about why it can be useful for writing center work. I will then introduce two IRB-approved research projects that Josh, Amy, and I have been working on over the last few years. Both projects involved coding the written artifacts generated during Writing Center programs. In Josh's case, the written artifacts were tutor comments left in asynchronous online tutoring. In Amy's case, the artifacts were the written feedback that graduate students left on their peers' documents as part of a writing group. After I have shared the methods and participants for both projects, Josh and Amy will spend the remainder of the time talking about the data from their respective projects. In particular, they will highlight how this sort of data can be used to tailor writing center practice to better train tutors and to better meet the needs of writers.

Slide 3:

So . . . Why code the comments that writers or tutors leave behind on documents they are reviewing?

In the writing center, we have theory: prioritize the global, be non-directive, ask questions, interact with NS and NNS in the same way. Tutors learn theory in a number of ways: tutoring practicum courses, staff meetings, Writing Center literature. We then expect them to put what they've learned into practice, to actually focus on the global rather than the local, to ask more questions than they give commands, to put into play the actual skills suggested by the theory. Ideally, theory and practice should be in perfect balance—the theory informing the practice, the practice mirroring the theory.

But how often do we examine whether what we **say** we do is the same as what we **actually** do in practice? How often do we examine our practice to see whether it matches what we preach? If you are like me, in the busy world of a writing center, days, weeks, months can go by without the time to do any serious, in-depth observation. And, once we do find time to observe, how do we “observe” sessions that have no in-person component? Sessions that result in written artifacts rather than interpersonal interactions? It is easy enough to simply read over transcripts of sessions or the comment bubbles tutors or writing group members leave on a document. It is harder to do so with a precise focus so that potential areas for improvement are spotted or so that disconnects between the theory and practice stand out.

One method for improving our perusal of written artifacts is coding them. I should probably note at this point that coding comments is in many ways an extremely tedious process and is not for the faint of heart. However, it results in a very detailed, clear picture of exactly what is going on in a session. Is the tutor actually commenting in similar ways to both NSs and NNSs? Are writing group members able to focus on content and discipline-specific material, or do they focus exclusively on APA style formatting issues or correcting grammar mistakes? Tinkering with comments in this manner (in essence spying on the material left behind by tutors and writing group members) offers insight into how we can tailor practice—how we can better train tutors to work in asynchronous online environments or how we can teach writing group members methods for improving the content and focus of their peer feedback.

Slide 4:

Now that we’ve considered the “why” of coding, let’s look at some methods for doing so. Our Asynchronous Project collected two forms of data: the comments that tutors left on papers during asynchronous tutoring sessions and transcriptions of interviews with those same tutors. We use the WOnline software for asynchronous tutoring. With this software, writers upload a draft of their paper (preferably in a word processing format), and tutors download the paper and add comments by using marginal bubbles and inserting summative comments at the beginning or end. The tutor then uploads the paper to the system, and writers can access it there.

The written comments left by tutors on writer’s papers were coded using a coding system that was originally developed and used at Josh’s former institution. You can access a full account of that in the *Learning Assistance Review* article cited on the slide and the handout.¹ This system involved coding both marginal comment bubbles and summative end comments using NVivo. Because we used NVivo, we could also test inter-rater reliability, which we measured by calculating both the Kappa Coefficient and the percentage agreement for 10 of our 183 coded documents. I’ll say a little bit more about this in a moment.

The coding system for this project looked at three things: the TYPE, FOCUS, and MODE for each independent clause a tutor wrote. Each independent clause thus had 3 codes attached to it.

¹ Weirick, J., Davis, T., & Lawson, D. (2017). Writer L1/L2 Status and Asynchronous Online Writing Center Feedback: Consultant Response Patterns. *Learning Assistance Review (TLAR)*, 22(2), 9-38.

This also means that any one comment bubble might have more than one set of 3 codes attached to it, depending on how many independent clauses were in the single bubble.

TYPE of comment determines whether the tutor is addressing GLOBAL issues such as content, organization, and point of view, or SURFACE issues related to individual sentences, such as clarity, word choice, and sentence structure. In addition, Type allows for labeling a comment as META-TEXTUAL, meaning that it refers to something non-rhetorical, such as a description of what a tutor will do in a session or an explanation of why the tutor did not cover everything a writer asked for. Meta-textual comments received only that one code, as, unlike Global or Surface comments, they did not have a focus or mode.

The FOCUS of a comment explains what the comment is actually about: A Focus on CORRECTNESS means the comment is correcting an error, a Focus on ORGANIZATION means the comment is focusing on the order of pieces in relation to each other. In particular, such a comment might indicate that this order should be changed during revision. A Focus on CONTENT indicates that the comment is interacting with the content in some way, perhaps suggesting that content be added or removed. A Comment on STYLE might be addressing issues of CLARITY (such as sentence structure or word choice), issues of FORMAT (such as APA citation style), or issues of GENRE (such as level of formality).

And finally, the MODE of a comment looks at how the focus is communicated: as Advice, Questions, Explanation, Praise, Commands, Criticism, or Recasting. Mode also allows for a miscellaneous category in case the focus is communicated in a way that doesn't appear to fit with existing options. For instance, if tutor comments seemed to be merely pointing out things the writer was doing without either praising or criticizing, they got coded miscellaneous.

The **interview** portion of the study consisted of one-on-one interviews with the tutors whose comments were coded. Tutors were asked about their WORK HABITS—how often they tutor online, how many comments they give in one session, and how long it takes them to comment on one submission. They were asked about their THOUGHT PROCESSES—how they decide what to comment on or how to phrase comments on an identified issue, what common strategies they use, or what they do if they don't know how to comment. And they were asked about their FEELINGS ABOUT ONLINE TUTORING—how much they enjoy it.

Slide 5:

One of the benefits of using a coding software like Nvivo is that, with the click of a button, it compares how multiple raters are coding the same document. I'm not going to pretend to be a statistician, but Josh and I did use this feature to check that we were coding comments similarly. According to both measures of inter-rater reliability, our separate coding of the same documents was reliable across all coding nodes.

The majority of our Kappa coefficients fell into the excellent category, with a small number in the fair to good. None were lower than that.

For the vast majority of coding nodes, our percentage of agreement was over 90%. And if you have further questions of the statistical nature, you'll have to check with us at a later point. Preferably much later.

Slide 6:

Before we move on to our second set of methods, I wanted to give you some idea of what this coding method looks like in the real world. The tutor's sentences on the left were part of a single comment, but the coding separates them into two sets of codes. Notice that both are Global, but one is focusing on content and one on organization, and one takes the form of praise while the other is a command. Notice also that technically the first piece is 2 independent clauses and would normally be coded separately.

This sample also demonstrates why we didn't just attempt to code the entire comment bubble as a single comment—there is just too much variation within a single bubble. And two sentences in a bubble is actually a fairly short comment. Many of the comments we worked with included 3 to 5 sentences per bubble. I should note that some comments are more challenging to code than others, and Josh will go into that more when he talks about his data.

Slide 7:

Our second project involved coding the comments that graduate writing group members left on their peer's documents. We became particularly interested in studying these during a semester when the writing group suddenly had a natural split between advanced members (who tended to dominate the conversation) and brand-new graduate students (who seemed to participate very little in the conversation). There were no middle-year students. We hoped to learn something about the learning trajectories of writing group attendees that would inform our decisions about how to structure the group and about what to overtly teach as part of the meetings.

For this project, we coded comments manually—we printed out the list of comment bubbles from the Word document and marked them with a pen. We also checked inter-rater reliability manually, using a compare and discuss method. In this project, comments could receive more than one code because we did not individually code each independent clause but instead coded each comment as a unit. I should note here, that the manual method worked because we have a much more limited data set than in the Asynchronous Tutoring project.

We coded each document with two different methods. Method 1 looked at the STYLE of feedback—was it corrective, directive, interactive, or evaluative? In other words, was the reader simply making corrections versus directing the writer to make them? Was the reader talking to the writer about the text and the possible corrections? Or was the reader offering an evaluation of the text? Method 2 looked at the FOCUS of the feedback. Did the reader simply insert or delete text? Did he or she focus on discipline-specific issues or organization? Or was the focus at the sentence level, with vocabulary issues, grammar, or clarity?

You might notice that there was some overlap between these two methods, but they do not map precisely onto each other. For instance, Corrective Feedback in Method 1 tended to correspond to Deletions and Insertions in Method 2, but readers could also “correct” by simply recasting in a comment bubble instead, which would not be coded as a deletion/insertion pair, nor would it be coded as directive because the actual comment did not tell the writer “do this.”

Slide 8:

The sample coded comment demonstrates what this coding method looks like in real life. The commenter says, “I would lead with this thought. The logic here should be that **DEPLOYMENT** [. . . and there you have the missing soldiers!] that deployment poses a threat to the attachment system, thereby causing family members to experience anxiety and threat regarding safety and comfort.”

For Method 1: The “I would lead” phrasing results in the comment being coded Directive. The rest would be coded Interactive—the reader is talking to the writer about the material.
For Method 2: That first sentence is about the organization of the paragraph (as seen through the use of the word “lead”, which implies moving the existing thought), and the rest is discipline-specific content.

Slide 9:

As you can see, our two coding systems are quite different, but there are some important similarities.

First, each method was developed after reading over only a few samples. In essence, our coding systems were a mix of a priori codes and emergent codes. In other words, while we did look over some samples to see what sorts of comments were being left and how they might be grouped into categories (emergent codes), we also started with some idea of what we were looking for based on writing center values (the distinction between global and surface concerns, for instance), and we decided what should be coded for (a priori) and then applied the codes to existing text, rather than going through all of the texts first seeking common words and determining after the fact what the codes should be.

Second, the emergent/a priori hybrid list received some refinement as we initially applied the list. For instance, we had to determine things like how “awk” as a comment should be coded or whether the addition of a question mark to a comment was enough to move it from directive to interactive. Also, since no two people comment in exactly the same way, we had to periodically ask how a particular turn of phrase should be coded.

Third, this refinement led to the development of a sort of coding scheme, a code book if you will (the handouts you have), where we listed the codes and their meanings, along with examples or hints about word choice that would lead to applying a particular code. Developing a codebook helps with inter-rater reliability and allows coders to rely less on their own interpretation and more on how particular words are grouped with particular codes. So, for

example, use of “should” versus “might” would determine whether a comment was directive or interactive, in the graduate writing example. Such a codebook also allows other researchers to replicate the study, if they wish.

Fourth, and perhaps most important, the codes were based on writing center values rather than on writer values. Hence, there is emphasis on the differences between global and local, between style and correctness, between criticism and praise. Any coding method that starts with some level of a priori codes will be based on the values of the researcher rather than the values of the subjects. This is not necessarily a problem—just something to keep in mind as you undertake a project like this.

Slide 10:

Before turning things over to Josh and Amy, I wanted to give you some idea of the participants in each study.

The Asynchronous Tutoring study included two, native-English-speaking, 1st year, graduate tutors, who we named Christen and Ellen. Christen conducted 101 online sessions during the study, and Ellen conducted 82 sessions, for a total of 183 sessions we had to code. Between the two of them, they worked with a total of 94 writers. 38 were Native Speakers & 56 were Non-Native Speakers of English. Among the writers, there were 38 undergrads and 56 who were graduate students or higher (there were a small number of visiting scholars, post-docs, and faculty included in the sample). 17 of the writers had 2 sessions and a further 15 had more than 2 sessions, with 1 writer having a total of 11 sessions.

The Graduate Writing Group study focused on 2 writers from a graduate writing group conducted by Writing Lab staff for a research institute in the social sciences. This was an initial test run of the data to see how fruitful it would be for our purpose of improving the writing group, so we limited it to the two writers for whom we had the most data at different stages. We named the two writers Mary and Lisa. Mary was an upper-level graduate student—by the end of the study, she finished her program and graduated. She was also experienced in the writing group, having been present from its beginning in Spring 2014. During the time in which we coded comments, Lisa was a 1st year grad student and, consequently, in her first year as a writing group member.

Now I'll turn things over to Josh and then Amy, so that you can hear what we learned about our programs, our writers, and our tutors.

Slide 11: Josh

Hi everyone, as Vicki said, my name is Josh and I will be presenting some preliminary findings from our investigation of Writing Lab tutors' patterns of response in asynchronous online Writing Lab sessions. I will first provide a brief description of the impetus for this project, followed by the general results of our study. I will then describe three applications of the method presented here that represent three different ways of looking at this data set. My aim in doing this is to demonstrate the flexibility of this

approach, and show how looking at the same data set in different ways can answer different research questions and have different implications for tutor training. Finally, I will discuss the challenges inherent in developing a coding system for asynchronous online comments. Since Vicki has already discussed the coding system, participants, and data collection, I will not discuss the methods of this study in great detail; however, I would be happy to answer questions about the methods should they arise.

Slide 12:

While tutors in the Writing Lab receive training for face-to-face sessions, they receive minimal formal training for asynchronous online tutoring. Although a more formal training program is being developed, currently training consists of a meeting with the Associate Director of the Writing Lab and a number of suggested readings. Since online tutoring is very new for us and the training is minimal, we are in a position to ask ourselves a couple of questions: In what ways would tutors benefit from training for asynchronous online tutoring? What sort of training should we be providing, and what topics should be covered? In order to answer those questions, we need to know what tutors are currently doing in asynchronous online sessions. In other words, the main motivation for our study was to answer the question “what are Writing Lab tutors actually *doing* when they leave comments in online sessions?” We also wanted to answer that question in a way that was a bit less impressionistic than the answer we might get if we simply chose a number of student papers and read the comments that tutors left. We wanted a way to systematically examine these comments, and we wanted to organize the data in such a way that it could be manipulated to answer more specific questions. We believe our coding system meets both of those needs.

Slide 13:

First, let’s take a look at the overall results of the coding. This slide shows each code, grouped by macro-category (i.e., Type, Focus, and Mode) for both of our tutors, Christen and Ellen. For each pair of bars, Christen is represented on the left in blue while Ellen is represented on the right in orange. Note that this chart shows the counts for each code. In other words, this is simply the ‘raw data’ from the coding.

For Type, Surface level feedback was more frequent than Global for both tutors; however, Christen appeared to be more balanced between Global (1183) and Surface (1400) compared to Ellen, who seemed to greatly favor Surface (1528) over Global (843) level feedback. The most frequently occurring Focus for Christen was Style (1018), while the most frequently occurring Focus for Ellen was Content (850). The most considerable difference in Focus seems to be in Correctness, which was more frequently a Focus of Ellen’s comments (735) compared to Christen’s comments (445). Finally, the most common Mode for Ellen was Explanation (516), while the most common Mode for Christen was Questions (616; though this may not be entirely accurate—more on that later).

Already, we can observe some interesting patterns that have implications for tutor training. For instance, there appears to be a difference between the two tutors in the distribution of Surface and Global level feedback: Christen seems to be fairly balanced between Global and Surface, while Ellen appears to favor Surface level feedback over Global level feedback. In terms of tutor training, this might suggest that we should expect differences in this distribution between individual tutors. In other words, we should not count on tutors to universally favor Surface over Global, or expect that they should be balanced between these two comment Types. Instead, this distribution may vary by tutor, and we may want to explore the reasons for that variance in order to provide more precise training.

Slide 14:

In addition to the coded data, the tutors participated in interviews where they were asked questions about their commenting patterns. For the most part, the answers that tutors gave in interviews complemented the patterns observed in the coded data, and sometimes provided some tentative explanations for those patterns.

For example, consider the pattern of comment Type observed on the previous slide: Christen tended to strike a balance between Surface and Global level feedback, while Ellen tended to give more Surface level feedback than Global. The answers that tutors gave in interviews provided a possible explanation for why this might be. In response to the question **“What would you say are the most common issues that you address in online submissions?”**, Ellen unambiguously indicated that she tends to focus on Surface level issues like grammar. Specifically, Ellen said that she focuses on grammar because writers request help with this area when they fill out the appointment form. On the other hand, Christen’s response to this question is framed very differently: Christen said that “the most common request is grammar,” but that she makes a conscious effort to also give feedback on Global level concerns.

Our tutors’ responses to this question largely complement the patterns in their coded data. They also represent two distinct strategies for determining which feedback areas tutors should spend their time commenting on. On the one hand, Ellen makes an effort to comment on the areas that the writer requests help with: if the writer requests help with grammar, Ellen will give more feedback on Surface level issues. On the other hand, Christen tends to let her perception of the needs of the paper determine which areas she should comment on: if Christen believes that the writer will benefit from additional Global level feedback, she will make an effort to leave Global comments regardless of whether the writer explicitly requested them.

Both of these strategies are valid, and it might be worth exploring them with tutors in our tutor training. For example, we could ask our tutors to reflect on which strategy they tend to prefer, and consider situations where one strategy might be more appropriate than another. The patterns observed in our data so far seem to indicate that tutors will differ on the distribution of Global and Surface level feedback, and may also differ in their motivations for commenting a certain way. In order to help our tutors become aware of their commenting patterns and help them be able to articulate what motivates them, we may want to consider addressing these differences explicitly in our tutor training.

Slide 15:

If we look again at the overall results, we can also see a noticeable difference in the frequency that our tutors asked questions in their feedback. Christen seemed to leave many more comments including Questions (616) compared to Ellen (204). In fact, Questions were Christen’s most frequently-used mode, used even more often than other common modes like Explanation and Qualified Command.

Slide 16:

The pattern of Questions in Christen's coded data was complemented by her interview answers. In response to the question "**When you've identified something to comment on, how do you decide what to say?**", Christen reported that her initial response was to ask a question. She gave several examples of questions that she might include in her comments ("I'm not following you here, what is this?" "What do you mean here?" "Is this accurate?"), and she went on to say that her frequent use of questions is a reflex of the tutoring strategies she uses in face-to-face sessions.

On the surface, Christen's use of questions seems highly desirable from a tutor training perspective. After all, tutors are trained to use strategies like Socratic questioning to help foster a collaborative environment in the tutorial and to ensure that the tutorial is writer centered. However, closer investigation of Christen's Questions indicated that many of them were questions about Surface level issues that, perhaps, could have been better addressed by simply offering a correction. For example, questions like "should there be a comma here?" and "should this be capitalized?" were quite frequent.

In terms of tutor training, we may want to consider whether these are the kinds of questions we want tutors to be asking. Asking questions is a cornerstone of writing center pedagogy and is especially important in online environments that tend to prompt more directive comments. But it could be argued that questions like "should there be a comma here?" are not truly questions, but are instead corrections 'dressed up' as questions in order to adhere to best tutoring practices. Additionally, this kind of feedback might be confusing to non-native English speaking writers, who are being asked to navigate through the many layers of the pragmatics of English question-asking in order to determine whether these kinds of comments are actual requests for clarification or whether they are something else. In our tutor training program, we may want to ask our tutors to think critically about the best practices we ask them to adhere to. This sort of critical thought will help to ensure that they are not following best practices arbitrarily and unreflectively. For example, this could include a discussion of which issues are well suited to being addressed through questions and which are better suited for other modes.

Slide 17:

While the general results reveal some interesting patterns and have some implications for tutor training, we need to ask some more specific questions about our tutors' comments to fully determine which areas our tutor training program should focus on. To that end, I will now transition to discussing additional ways to manipulate these data to answer three more specific questions: Do tutors' patterns of response differ based on writer language background? Are summative comments different than side comments? And do tutors' patterns of response change over time?

Slide 18:

First, I will discuss differences in tutor's patterns of response in comments to writers whose first language is English (L1) and writers whose first language is not English (L2).

This slide shows the two Type codes, Global and Surface, for both Christen and Ellen. Feedback to L2 writers is shown on the left side of each pair of bars in blue, and feedback to L1 writers is shown on the right side in orange. These bars represent each code as a proportion of the macro-category. For example, in Ellen's feedback to L1 writers, 62% of Modes were Global, while about 38% of Modes were Surface. In all of the charts to follow, each code will be shown as a proportion of its macro-category.

For both tutors, L1 writers tended to receive more Global level feedback, while L2 writers tended to receive more Surface level feedback. For Ellen, approximately 62% of her feedback to L1 writers was Global, while about 38% was Surface. For L2 writers, the reverse is true: about 24% of Ellen's feedback to L2 writers was Globally oriented, while about 76% was Surface oriented. Christen's feedback shows a similar pattern: about 65% of Christen's feedback to L1 writers was Global, while about 35% was Surface. For L2 writers about 40% of Christen's feedback was Global, while about 60% was Surface. In general, this pattern indicates that, for both tutors, L1 writers tended to receive more feedback on Global level issues, while L2 writers tended to receive more feedback on Surface level issues.

Slide 19:

Shifting to Focus, we see a pattern that largely complements the pattern observed in the coding for comment Type. This graph shows the Focus of feedback Ellen gave to L1 and L2 writers. L1 writers tended to receive more feedback on content (59%), while L2 writers tended to receive more feedback on correctness (41%). These results seem to suggest that L1 writers receive more feedback on Content, while L2 writers tend to receive more feedback on Correctness.

(Note: the pattern described above held for both tutors, so only Ellen's feedback is shown here).

Slide 20:

Transitioning to Mode, we begin to see some differences in our tutor's patterns of response. Looking at the distribution of Modes in Ellen's feedback, L2 writers appear to receive more feedback that includes Explanation (22%) compared to L1 writers (15%).

Slide 21:

Looking at the distribution of Modes in Christen's feedback, we see the opposite pattern. While Ellen tended to include more Explanation in feedback to L2 writers, Christen more frequently included Explanation in feedback to L1 writers.

Slide 22:

Again, the patterns observed in the coding were also supported in the answers the tutors gave in their interviews. Recall that both tutors tended to give L2 writers feedback that was more Surface-oriented and more focused on issues of Correctness. In response to the question "**What would you say are the most common issues that you address in online submissions?**", Ellen said that, in her experience, international students often request help with grammar. This fits well with the pattern observed in the coded data, which suggested that L2 writers tend to receive more feedback on Correctness.

Based on the patterns observed in the data, there are several implications for tutor training regarding how tutors interact with L2 writers in asynchronous online sessions. It might not be entirely unexpected that L2 writers appear to receive more Surface level feedback and more feedback on Correctness. After all, it is entirely possible that L2 writers receive more of these kinds of feedback because L2 writers tend to request help with grammar. However, both tutors also gave L1 writers more feedback on Content compared to L2 writers. This might prompt us to consider whether the increased frequency of comments on grammar for L2 writers is coming at the expense of comments on Content. In other

words, we may be prompted to ask ourselves whether these tutors are paying a bit too much attention to grammar issues and too little attention to Content issues in papers by L2 writers. In our tutor training, we might consider working with our tutors to help them develop an awareness of Global level issues in L2 writing and strategies for effectively commenting on those issues. At the very least, we may want to ask our tutors to reflect on their patterns of response in sessions with L2 writers so that they are aware of the danger of getting too caught up in Surface level issues.

Slide 23:

The second application I would like to discuss is an analysis of summative vs. 'side' comments. In the Writing Lab, tutors are encouraged to end each asynchronous online session with a closing, summative comment. This comment generally recaps the areas the tutor has commented on, and some tutors also use it to give some holistic feedback on the whole paper. Summative comments are generally (but not always) located at the beginning of the document, and can be anywhere from a few lines to several paragraphs long. By 'side' comments I mean the Microsoft Word comment bubbles that are written by tutors as they read through a document.

This slide shows the distribution of comment Type for summative and side comments for each tutor. For each pair of bars summative comments are shown on the left in blue while side comments are shown on the right in orange. For both tutors, summative comments tended to be more Globally oriented, while side comments tended to be more Surface oriented.

Slide 24:

Turning to Focus, for Christen summative comments tended to be more focused on Content (46%) and Organization (15%) while side comments tended to focus on Correctness (19%) and Style (42%). The results for the Focus of Ellen's summative and side comments (not shown) also show this pattern.

Slide 25:

In terms of Mode, Ellen's summative comments show noticeably more Praise (23%) compared to her side comments (5%). Side comments tended to have more feedback that included Qualified Commands (23%) and Questions (10%).

Slide 26:

Like Ellen, Christen's summative comments tended to include more Praise (11%) compared to her side comments (4%). Unlike Ellen, Christen's summative comments tended to include **more** Explanation (22%) compared to her side comments (14%). Christen's side comments tended to include more Questions (27%) compared to her summative comments (19%).

Slide 27:

The patterns observed in the coded data were supported by the answers tutors gave in interviews. Recall that summative comments tended to be more Globally oriented and tended to focus more on areas like Organization and Content, while side comments tended to be more Surface oriented and focused more on Correctness and Style. In response to the question "**When you've identified something that you need to comment on, how do you decide what to say?**", Ellen indicated that she

felt more comfortable addressing “larger order concerns” like organization in a summative comment. Christen said that if she was commenting on grammar, she tended to leave side comments as she read. In other words, Christen generally addressed grammar using side comments.

In summary, summative comments tended to be more Globally oriented, tended to focus on higher order concerns like Organization and Content, and tended to include more Praise for the writer. In terms of tutor training, this seems like a very desirable result. This might prompt us to encourage, or even require, tutors to leave summative comments as part of our asynchronous online training program.

Slide 28:

The final application I would like to discuss is an analysis of change over time. In order to analyze change over time, we selected 20 asynchronous online sessions conducted by each tutor with L1 writers: 10 sessions from ‘early’ in the tutor’s career, and 10 from ‘late’ in the tutor’s career. Comments from those two samples were then analyzed and the results compared. This slide shows the distribution of comment Type in Ellen’s early and late sessions. For each pair of bars the early comments are shown on the left in blue while the late sessions are shown on the right in orange. As this graph shows, the distribution of Type in Ellen’s early and late comments is nearly identical, indicating virtually no change over time. The same pattern was found in the analysis of Christen’s comments, and in the analysis of Focus for both tutors.

One explanation for these results is that there was genuinely no change over time; however, there are several other possible explanations. For example, it was quite difficult to find 10 sessions with L1 writers clustered near each end of the timeline; instead, the 20 sessions were more or less evenly distributed across each tutor’s career, and this even distribution could have obscured any change-over-time effects that might have otherwise been observed.

While no changes over time were observed in this analysis, it is easy to see how such results could be valuable for tutor training. For instance, having information about how tutors’ patterns of response change over time could help us identify common pitfalls that new tutors encounter, which could then be addressed more thoroughly in training. Additionally, if we were able to identify how tutors’ patterns of response change as they become more experienced, it might be possible to design our training in such a way as to help tutors progress more quickly and achieve ‘veteran’ patterns of response earlier than their forbearers. While this analysis was ultimately unenlightening due to the difficulty in finding true early and late sessions with L1 writers, change over time seems a valuable area for future research.

Slide 29:

While the analysis of change over time on the whole did not reveal many significant patterns, one clear pattern was observed. This slide shows the distribution of Mode in Christen’s early and late comments. One noticeable difference is in Explanation: in Christen’s early sessions, Explanation only made up 9% of all Modes used. In late comments, Explanation made up 29% of all modes used, a 20% increase. This result could indicate that Christen became more comfortable offering explanations for various issues as she became more experienced. If this is in fact the case, it suggests that we may want to work with tutors in our training program to help them develop go-to explanations for frequently occurring issues. These can be incorporated into their comments until they become comfortable giving explanations on

the fly. While the increase in Explanation over time should be treated as extremely tentative given the issues described above, Explanation seems a clear area to look at in future work.

Slide 30:

In summary, this presentation has discussed three applications of our coding method, each of which helped us to understand what our tutors' patterns of response looked like in asynchronous online sessions. Based on these patterns, we were able to see some implications for our tutor training. For instance, the general results suggested that some tutors tend to maintain a balance between Global and Surface level feedback while others tend to favor one over the other. We speculated that one reason for this might be that multiple strategies exist for determining which feedback areas a tutor should focus on in a paper. We indicated that our tutor training may want to incorporate a discussion of these strategies. We observed that both tutors tended to give L2 writers feedback that was more Surface oriented and more focused on Correctness, while L1 writers received feedback that was more Globally oriented and more focused on Content. This finding prompted us to consider whether or not it was desirable for L2 writers to receive feedback on Correctness at the expense of feedback on higher order concerns, and this indicated that we may want to help tutors develop strategies for effectively mixing in Global level feedback in sessions with L2 writers. We also found that summative comments tended to be more Globally oriented, tended to focus on Content and Organization, and tended to include more Praise. This seems like a desirable result, which might indicate that we want to encourage tutors to write summative comments in our training program. Finally, while the change over time analysis did not lead to any conclusive results, we considered some of the ways that the results of this application could be valuable when developing a tutor training program, and we identified one specific area, change in Explanation over time, which might make for a fruitful subject of future work. These are just a few of the many ways that this type of analysis can help us understand the kind of feedback tutors are giving in asynchronous online tutoring sessions and how those findings can help us tailor our tutor training programs.

Before concluding, I would like to briefly discuss some of the challenges inherent in coding online comments. This slide shows two comments we encountered in our study and how they were coded by both members of our research team (Vicki and Josh). For the first comment, Josh coded it as being focused on Content, while Vicki coded it as being focused on Organization. Either of these codes can be justified depending on how the researcher interprets the comment. For example, this comment indicates that the writer should add transitions. A transition could consist of a single sentence added to a single paragraph, or a number of sentences added to a paragraph, or a number of sentences added to each paragraph throughout the paper. Depending on how many transitions are needed and where, this comment could be prompting the writer to make a fairly substantial revision. Since comments that prompt the writer to include additional content are considered to have a Content Focus, a suggestion to add transitions could be considered a Content comment. On the other hand, transitions are an organizational device, and one could easily justify coding the comment for Organization as a result.

In the second comment, Josh coded the Focus as Correctness, while Vicki coded it as Style-Clarity. This comment is praising the writer on punctuation use. The aspect of the praise that is salient to the researcher can affect how the comment is coded. For example, if a researcher keys in on "clear," one could justify coding the focus of this comment as Style-Clarity, which would assume that the tutor was complementing the writer on how their comma use contributed to the Clarity of the whole sentence.

On the other hand, comma use is generally considered to be a matter of Correctness, since the rules for most comma usages are well defined and widely known. To say that the comma use is “appropriate” seems to praise the writer on following the rule, which would justify coding the Focus of this comment as Correctness. Since it is not possible for the researcher to establish the intention of the tutor for every comment, ambiguities like these are ubiquitous. The functions of comments and the factors that motivate tutors to leave a particular kind of comment in a particular situation are very complex, and any attempt to characterize comments using only a few features will necessarily require the researcher to make choices in interpretation. Researchers can ensure that members of the research team are making the same choices in interpretation for the same comment with reasonable consistency by making sure the coding system is well defined, and by setting aside a good amount of time for norming. While challenging, our experience shows that performing this kind of analysis is worth the daunting task of coding and can lead to valuable insights.

With that, I will hand it over to Amy for the next section of our presentation.

Slide 31: Amy

As Vicki mentioned in the introduction, we are working with a graduate student writing group, composed of graduate students in the Health and Human Sciences who primarily work on research for military families. So, we are quite literally tinkering with and tailoring papers about US soldiers. Now for the spying. We selected two students, Mary, a returning writing group member and an advanced graduate student (dissertating), and Lisa, a first-year graduate student new to writing group. We noticed that the first-year students tended to be a little quieter during group discussion, and we wanted to investigate if their quietness reflected having fewer substantive comments on papers, or perhaps if there were other factors (e.g., pecking order, shyness, confidence ability issues, etc.). We are interested in looking at trajectories, curious about how both style and focus of comments may change over time. Specifically, we chose two different methods (style *and* focus) because any of the content could be done in any of the styles. Our tentative hypothesis, based on Writing Center values, is that over time we would expect to see writers improve, as measured by numbers of interactive comments (style) and higher order concern comments (focus). We’re interested in the implications this can have for establishing and supporting writing groups and the role that writing centers can play in this.

Slide 32:

Let’s start with Mary, our experienced writing group member, and look at method 2 (see the handout)—coding for focus. Each “dot” on this graph represents one paper Mary commented on. She commented on 14 papers across four semesters. You’ll notice that for the very first paper (on the left side of the graph), there are almost no “organization” (yellow line) comments and a high number of “insertions” (orange line). Insertions tend to be direct grammar fixes (anything from spelling to commas). Of the 60 comments left on that first paper, over half of the comments were insertions/deletions. For the first few papers, in fact, insertions and deletions tend to be quite high, despite the fact that the papers’ authors asked for help with

tying the content back to theory, determining how many examples to include, organization, and transitions. Almost all of the deletions/insertions were grammar-related.

However, we see a “flip point” by the fifth paper she commented on, where the lines on the graph invert: insertions/deletions (orange and blue) decrease, and organization and discipline-specific increase (yellow and black). This trend remains relatively stable for the rest of Mary’s time in the writing group. We’re interested in this kind of trajectory because Mary’s data suggests that, throughout her time in writing group, she became better able to provide higher order comments on organization, content, data handling, etc., and thus could help (or felt she could help) her peers more.

Slide 33:

We also see from Mary’s data that certain types of comments tend to mirror each other. Looking at the graph on the left, perhaps unsurprisingly, insertions and deletions tend to go together, as when someone deletes the comma and adds a period. Looking at the graph on the right, it seems too that discipline-specific, organization, and sentence structure comments all tend to rise and fall with one another.

Slide 34:

Similarly, we can see that deletions and discipline-specific comments oppose each other, such that when Mary makes specific corrections in a text, she makes fewer marginal comments on content and vice versa. For example, looking at the 3rd paper Mary commented on here (paper 3), 40% of the comments are deletions (blue line), compared to just over 10% as discipline-specific comments (black line). On the other hand, by the 9th time Mary commented (paper 15), around 14% of the comments are deletions (blue line) while over 50% of the comments are discipline-specific (black line).

This data is also quite helpful as we think about writing group facilitation and how to help graduate students become better commentators. It seems that when they focus too closely on text-based corrections or other lower order concerns, they may ignore higher order concerns.

Slide 35:

For Lisa, our new writing group member, the data looks a little different. She commented on six papers throughout her first year. Similar to Mary’s first year in writing group, there tends to be more insertions and deletions than any other category. Throughout the entire year, Lisa made *no* organizational comments. Also similar to Mary’s data, deletions and insertions tend to mirror each other over time, and Higher Order Concerns and Lower Order Concerns tend to oppose one another.

We do have one outlier, and that is for “discipline-specific” (black line) on the 3rd paper she commented on (paper 15). We were surprised to see a sudden spike here, when the data was otherwise fairly constant and showed Lisa leaving fewer higher order concerns than Mary

overall. We returned to the paper to look at the specific comments, and these comments were almost exclusively on APA content/formatting (73 separate comments on formatting—italicizing, links to the APA handbook, etc.). On the handout, you'll see that APA formatting is classified as "discipline-specific," which are usually higher order concerns.² Since so many of these comments, though, were on italicizing, which is really just formatting, these comments don't actually address higher order concerns, but rather lower order concerns.

With Mary, by the 5th paper she commented on (paper 7) we saw a switch to more global concerns, but we don't see that same kind of movement here. Lisa was new to the academic program, so it's possible she felt unsure about content or hesitant to express opinions/ideas around more advanced students. Anecdotally, the group did look to her as an APA expert, and most of her verbal comments during writing group meetings were also about format. As we think about future writing group development, it may be very useful to work closely with new graduate students about what to comment on and also to work on confidence boosting. It's important to think about group dynamics and how newer students can learn from more advanced students (without the more advanced students feeling like they must take on the brunt of the labor.)

Another element to consider is that we didn't initially teach writing group members how to comment. The data here reflect what the writer was previously taught to do, what the writer thought was correct, or what was a natural impulse. A lot of Lisa's comments were on lower order concerns. As we think about facilitating writing groups, whether to teach/train (and what to teach/train) prior to joining a group may be something to keep in mind, especially if the writing group contains a wide range of experience. A few of our advanced students expressed frustration with the comments they received from new members, as the comments were less helpful than what they wanted or did not focus on what they wanted help with. One implication of the research is to think about the potential drawbacks of mixed-age grouping. There are advantages, of course, such as newer members learning from older members, but whether these advantages outweigh the disadvantage of unequal commenting should definitely be considered. When creating writing groups, both group composition and the amount and nature of comment training are important issues to discuss.

Slide 36:

More specifically, we can look at Year 1 for each writer and see some interesting trends, which allow us to draw possible implications for Writing Center work. We can't always control for number of years in graduate school (and Mary began participating in writing group later in grad school than Lisa did). However, we have the same number of papers from each writer over the

² Our original work with these methods was based on Mary and her cohort (i.e., more experienced graduate students and writing group members). For that group of writers, any reference to APA tended to fit with the category "handling literature within the disciplinary field" more generally, hence the inclusion of "style guide" issues under discipline-specific (HOC). Future research using this method might require an update to the method.

course of their first years, and we saw one student improve and one student remain relatively stagnant. You'll notice in particular the gaps I've highlighted here, as percentage-wise, there is a sizable difference (40% and 8% for Lisa, and only 35% to 15% for Mary), with Lisa relying almost exclusively on insertions/deletions. I should note that sometimes this is a sign of confidence—writers make direct changes to the text for elements they are sure about. These differences between early- and late-attending writers could help inform writing group practice in two ways: students should be invited to join writing groups early in their graduate career, and writing group facilitators may want to teach them about how to leave comments as an important early step. Insertions/deletions can be helpful when writers are in the end stage of a project, certainly, but there are other ways to share the same information for improving a document. Discussions about the differences in types of comments could then potentially help writers as they continue in graduate school. The type of work we're thinking about here—types of comments and how to give them—can have real long-term implications. If these graduate students become faculty, this type of work can help them work with their future students and move them beyond leaving the dreaded “awk” or “?” comments, which are often quite unhelpful (especially for multilingual writers). Our data here is also useful for thinking about instructional help that writing centers can offer because across the board “organization” is low, despite the fact that many of the writers asked for help on “flow” and the logic of their writing.

Mary did not receive specific instruction in how to comment, and she still improved, possibly from years in the program, seeing more advanced writing, growing in confidence, etc. Anecdotally, we did spend time teaching a new writing group member last year about comment types, and she's said she really thinks about how she words comments and has caught herself saying something one way and changing it to another. So perhaps in future research we'll have to compare her data with Mary's and Lisa's to determine the extent to which overtly teaching comment type has long-term effects on writers' abilities to offer improved feedback.

Slide 37:

We were also interested in the “discipline-specific” category, as we broke that category down into subsections to reflect the types of content that Mary and Lisa looked at. We found that Lisa (in blue), our first-year, struggled with commenting on higher-order concerns. Here, we break down paper 15, in which Lisa was quite high in “discipline-specific.” This proved to be our outlier, as I mentioned earlier. You can see that though she is high in what appears to be a “higher-order concern” category, her comments are on APA style, which is actually a lower-order concern, usually important for late-stage writing. In contrast, Mary is high in data handling, an important concern for readers to be able to follow a writer's argument.

(This is the 8th paper Mary commented on and the 3rd for Lisa. For Mary's 3rd paper, she had 42% data handling, 7% measures, 0% visual models, 36% APA style, 14% content, 0% other.)

Slide 38:

In order to compare how Mary handled discipline-specific comments with how Lisa did, we broke down Mary's "discipline-specific" comments for one paper and found a much greater percentage of the comments were in fact higher order concerns. We see "content" in yellow is at 58%, data handling (gray) at 34%, and APA style (blue) at just 8%. This is the 11th paper Mary had commented on during her time in writing group.

Slide 39:

Next, we also looked at results from coding Method 1: the style of the comment. This method examines *how* writing group members leave comments rather than what they say. We are interested in how styles of commenting can change over time, what that may indicate about writers' abilities, and what implications this has for Writing Centers who facilitate writing groups. So, again, corrective comments are when the editor made the change in the text itself; directive comments are when the editor provides specific instructions about what change to make in a side comment but does not make the actual change; interactive comments are when the editor makes suggestions, asks questions, offers clarification, suggests possibilities, etc.; and evaluative comments are praise/critique.

We can see here that Mary is high in corrective comments in the early papers, but over time they more or less decrease but never completely go away. This isn't to suggest that corrective comments are bad (sometimes it is much easier to just add the necessary comma). What we're interested in is the percent of comments compared to other types. The data show a rise over time in "interactive" comments, as demonstrated by the best fit line on the graph. What this suggests is that over time, this writer has developed in her commenting skills and over time engages more with the text, asks more questions, leaves ideas, or offers general commentary. These things tend to be hallmarks of good peer reviewing. Note that they are especially important in a peer-peer dynamic; advisor-advisee relationships, in contrast, may legitimately have more of a directive feel. On the other hand, anecdotally, we asked the writers about their style of commenting, and they said sometimes they leave questions when they're not sure about something. Our method for coding style placed questions squarely in the interactive category, but if writers use question marks to hedge what they perceive as directive comments, this might indicate that our method could bear more investigation to break down interactive comments. We generally find, overall, that an increase in percentage of Interactive (coupled with a decrease in corrective) tends to indicate improvement in commenting ability.

Slide 40:

If we look at Lisa's one year of data for Method 1, we tend to find that corrective comments dominate her peer reviewing. Looking at the comments themselves, we did find that a majority of these tend to be small types of errors (grammar, APA, spelling), which does make sense for corrective comments as editors don't normally rewrite paragraphs for the author. What we see across the year, though, is that there is no switch in style of comments or

measurable change. In Mary's case, by paper 5, we started to see change (though keep in mind that Mary and Lisa started attending writing group in different years of their academic program). Comparing Mary's and Lisa's data suggests a need for writers to be in writing group over time. A shorter amount of time, while it may help individual papers, does not seem to reflect sustained improvement in commenting or increased editing capabilities. For writing centers, then, when establishing or facilitating writing groups, our data suggests that the group's longevity and member retention should be part of the discussion, with writing centers seeking to develop long-term, stable groups rather than conducting "pop up writing groups."

Slide 41:

As with method 2, we can learn quite a bit by looking at trends with Method 1. We can see that "corrective" and "interactive" clearly oppose each other, such that when one is high the other is low. The data do not show that the writers can be high in both. This helps Writing Centers think through how we help writers. If they're focusing on lower order concerns, they're likely neglecting higher order concerns (rather than doing both).

Slide 42:

For comparison, we examined the 5 papers that *both* Mary and Lisa commented on to see the differences. We can see here a clear difference between Mary and Lisa's commenting styles. Even when Mary's corrective comments are high, Lisa's tend to be much higher. Paper 18 is particularly telling (note, this is *not* the one we identified as the outlier earlier). The author of this paper asked for help with "organization/flow and content," which do not lend themselves easily to a corrective style of feedback, yet Lisa's comments tended to be largely corrective. A high number of corrective could indicate confidence level issues; for instance, writers feel comfortable commenting only on what they are sure about (and therefore make direct changes). Writing Centers looking to facilitate writing groups ought to be aware of these tendencies among writers and consider how to help members learn to leave comments that engage with the text even if they're unsure of their own knowledge or ability. The data also help us see places where as Writing Center folks we can help writers think through how to address higher order concerns.

Slide 43:

Just for another quick comparison, here is the side-by-side chart for method 2 (content). We can see again, on paper 18, that the majority of Lisa's comments are deletions/insertions compared to Mary's substantive discipline-specific feedback.

Another revealing detail from this graph is the lack of "yellow," which is for "organization," across the board. For Lisa, it is non-existent, and even for Mary, our advanced writer, it is quite low. This could be one area for Writing Centers to help writers think through paper structure. Writing groups might offer writers language to address organizational issues they find in papers or an understanding of the types of questions that might be asked to clarify organization.

So, just as a brief wrap up, what we've found generally is that (not surprisingly) writing groups can be important and productive for writers. We see the most change over time, which could imply that just joining for a semester might not have the type of pay off that one would want from a writing group (even if in the short term a single paper improves.) Consistent with Writing Center ideals, we're interested in helping the writers develop, rather than becoming a one-stop-shop, and, as Mary's data shows, this happens across time. Our results indicate, too, the need for Writing Centers potentially to teach writing group members how to make comments and facilitate their development through activities, modeling, etc. We want to make writing groups sustainable and help members continue to develop, such that members help future grad students, and so on, and eventually their own students too.