Evolution of Instructor Response? Analysis of Five Years of Feedback to Students

Susan Lang, *The Ohio State University*

Structured Abstract

- **Background:** Research incorporating large data sets and data and text mining methodologies is making initial contributions to writing studies. In writing program administration (WPA) work, one could best characterize the body of publications as small but growing, led by such work as Moxley and Eubanks' 2015 "On Keeping Score: Instructors' vs. Students' Rubric Ratings of 46,689 Essays" and Arizona State University's Science of Learning & Educational Technology (SoLET) Lab. Given the information that large-scale textual analysis can provide, it seems incumbent on program administrators to explore ways to make regular and aggressive use of such opportunities to give both students and instructors more resources for learning and development. This project is one attempt to add to this corpus of work; the sample for the study consisted of 17,534 pieces of student writing representing 141,659 discrete comments on that writing, with 58,300 unique words out of over 8.25 million total words written. This data is used to examine trends in the program's instructor commentary over five years' time. By doing so, this study revisits a fundamental task of writing instruction—responding to student writing, and from the data's results considers how large writing programs with constant turnover of graduate teaching assistants (GTAs) might manage their ongoing instructor professional development and how those GTAs will improve their ability to teach and respond to writing.

- **Literature Review:** Researchers have attempted to unpack and understand the task of instructor commentary for several decades; the published literature demonstrates a complex and occasionally ambivalent relationship with this central task of writing instruction. Recent scholarship has moved from the small-scale studies long used by the field to implement large-scale examinations of the instruction occurring in writing programs.

- **Research questions:** Three questions guided the inquiry:

  1. Does the work of new instructors (MA1s) more closely resemble the lexicon of novice or experienced responders to student writing?

  2. How does the new instructors' work compare to that of more experienced (PHD1 or INS) instructors in the program throughout their time?

  3. How does their work evolve over a four-semester longitudinal time frame (as MA1 or MA2 experience levels) in the first-year writing program?

  Please note that the abbreviations used above and throughout the article to designate instructor experience levels are as follows: MA1 (first-year master's students); MA2 (second-year master's students); PHD1 (first-year doctoral students); INS (instructors—those with 3 or more years' experience teaching and who are not currently pursuing an additional degree—nearly all of these individuals held a Master's degree).

- **Methodology:** This study extends the work of Anson and Anson (2017) who first surveyed writing instructors and program administrators to create wordlists that survey respondents associated with "high-quality" and "novice" responses, and then examined a corpus of nearly 50,000 peer responses produced at a single university to learn to what extent instructors and student peers adopted this lexicon. Specifically, the study analyzes a corpus of instructor comments to students using the Anson and Anson wordlists associated with principled and novice commentary to see if new writing instructors align more closely with the concepts represented in either list during their first semester in the program. It then tracks four cohorts for evolution and change in their vocabulary of feedback over their next three semesters in the program; the study also compares the vocabulary used in their comments to that used by experienced instructors in the program over the same time.

- **Results:** The study found that from the outset, the new instructors (MA1) incorporated more of the principled response terms than the novice response terms. Overall, in comparing the MA1 instructors with the most experienced group (INS), the results reveal three important findings about the feedback of both MA1s and INSs in this program.

1. While there are some differences in commentary as seen via examination of the two lexicons, the differences are perhaps less than one might assume.

2. The cohorts do increase their use of the principled terms as they move through the two years' appointment in the program, but few of the increases demonstrate statistical significance.

3. Few of the terms from either the novice or principled lexicon, with the exception of terms that also appear in the assignment descriptions, what I label as "content terms," appear frequently in the overall corpus.

- **Discussion:** Based on the results, the instructors in this program had acquired a more consistent vocabulary, but not primarily one based on Anson and Anson's two lexicons—instead, the most frequent and commonly used terms seem to come from a more local "canon," that is, one based on the assignment descriptions and course outcomes. Regardless of whether the acquisition of a common vocabulary came from more global concepts or an assignment-based local canon, using common terms is something that Nancy Sommers (1982) saw as contributing to "thoughtful commentary" on student writing. As no one has previously studied how quickly new instructors acquire a professional vocabulary for responding to student writing, it is hard to know whether or not the results of this particular group of instructors would be considered "typical." However, it may well be that the context of this writing program contributed to a more accelerated acquisition.

- **Conclusions:** Working with the lexicons developed via Anson and Anson's survey is a useful starting point for understanding more of what our instructors actually do when responding to student writing, as well as for identifying critical differences in our instructors' comments. The lexicons, though, only provide us with a subset of expected (thus acceptable) terms included in commentary—terms that afford students the opportunity to act upon receiving them via revision or transfer.

- **Directions for Future Research:** Additional research is necessary to expand and refine the lexicons and their impact on student writing. One possibility is to return to the current data set to engage in additional lexical analysis of both the novice and principled lexicons as well as the overall frequency tables to understand how terms are used in the context of response by the various instructor groups. Differences in the application of the terms might help us understand why comments might be labeled as more or less helpful to writers. Another strategy is to examine the data in terms of markers of stance; finally, topic modeling could be used to locate more subtle differences in the

instructor comments that are not as easily identifiable with lexical analysis. Such examinations could serve as a baseline for broadening the study out to other sets of assignments and commentary, perhaps helping us build a set of threshold concepts for talking about writing with our students. Ultimately, it is important to replicate and expand Anson and Anson's survey to other stakeholder groups. As with much research on the teaching of writing, we default to the group most accessible to us—other writing professionals. Replicating this survey with other stakeholders—graduate teaching assistants, undergraduate students at both lower and upper division levels— could help us understand whether or not a gap exists in understanding what constitutes good feedback from the various stakeholders.

Keywords: first-year composition, instructor response, instructor training, writing analytics

1.0 Background

Research incorporating large data sets and data and text mining methodologies is now making initial contributions to writing studies. In writing program administration (WPA) work, one could best characterize the body of publications as small but growing, led by such work as Moxley and Eubanks' 2015 "On Keeping Score: Instructors' vs. Students' Rubric Ratings of 46,689 Essays" and Arizona State University's Science of Learning & Educational Technology (SoLET) Lab. Given the information that large-scale textual and data analysis can provide, it seems incumbent on program administrators to explore ways to make regular and aggressive use of such opportunities to give both students and instructors more resources for learning and development, as Lang and Baehr encouraged in 2012. This project is one attempt to add to this corpus of work; it uses data, nearly 150,000 discrete instructor comments on over 17,000 specific student writing assignments, to examine trends in the program's instructor commentary over five years' time. By doing so, this study revisits a fundamental task of writing instruction— responding to student writing, and from the data's results considers how large writing programs with constant turnover of graduate teaching assistants (GTAs) might manage their ongoing instructor professional development and how those GTAs will improve their ability to teach and respond to writing.

We know that commenting on student writing is critical—regardless of how faculty define "writing." Whether it be in first-year writing, writing in the disciplines, or graduate-level writing, or whether it be text-based or multimodal, the constant of writing instruction is that students receive a response. However, much of the research into instructor commentary on student writing has found that the reality does not measure up to the ideal, regardless of the instructor's experience level. Factors beyond the individual instructor's or even writing program's control may contribute to the less than ideal commentary, such as large class size or teaching load; it is still incumbent on WPAs, however, to seek ways to provide instructors with optimal preparation and feedback on their work when they are a part of our programs. Moss, Girard, and Greeno

(2008) assert that "assessment is (or should be) at least, in part, about professionals learning to support students' learning," and by extension, their own (p. 295). They contend that knowledge and actions of educators can be "shaped…by the assessment practices in which they engage" and ask whether we can design assessment systems to facilitate professionals' learning (p. 296). This project describes one way in which our first-year writing program provided ways in which to use the evidence generated from instruction, in this case, instructor commentary, to provide instructors with an opportunity to learn.[1]

The two types of primary artifacts (evidence) generated in the daily operations of a writing program are, of course, student writing and instructor commentary on that writing. Program administrators need ways to examine the state of and evolution of commentary as we train and supervise instructors, whether they be new master's students or professionals with years or decades of experience. And although administrators have been able to talk about and with their instructors about their own commentary, the agile and "deep assessment" that Dixon and Moxley (2013) refer to hasn't been possible, especially on a large scale, until large programs began responding to student writing in digital environments and creating ways to archive and access those responses

Despite the fact that administrative actions such as reading and evaluating instructor comments, especially with the use of technology, are often met with charges of micromanaging, infringing on the autonomy of instructors, and the like, I would contend that they can potentially benefit instructors, students, and the writing program overall. The following pages discuss one example of research devoted to advancement of opportunity to learn. This study extends the work of Anson and Anson (2017), who first surveyed writing instructors and program administrators to create wordlists that survey respondents associated with "high-quality" and "novice" responses, and then examined a corpus of nearly 50,000 peer responses produced at a single university to learn to what extent instructors and student peers adopted this lexicon. Specifically, it analyzes a corpus of instructor comments to students using the Anson and Anson wordlists associated with principled and novice commentary to see if new writing instructors align more closely with the concepts represented in either list during their first semester in the program. It then tracks four cohorts for evolution and change in their vocabulary of feedback over their next three semesters in the program; the study also compares the vocabulary used in their comments to that used by experienced instructors in the program over the same time. Doing so should reveal any significant patterns in where our new instructors start in their commentary and (potentially) how quickly they acquire the vocabulary of more experienced instructors.

---

[1] As Pullin (2008) notes, "if we want to understand learning and OTL [Opportunity to Learn], we must understand the relationship between learners and their learning environment. Learning is a continual and dynamic activity…for both students and educators…and each member in any learning community constructs (and reconstructs) an identity as a learner during interactions with others and with the learning environment" (p. 337).

2.0 Literature Review

Researchers have attempted to unpack and understand the task of instructor commentary for several decades; the published literature (e.g., Anson, 1989; Anson 2000; Brannon & Knoblauch, 1982; Connors & Lunsford, 1993; Keh, 1990; Sommers, 1982; Straub, 1996, 1996a, 2006; Straub & Lunsford, 1995; White, 2016) demonstrates a complex and occasionally ambivalent relationship with this central task of writing instruction. Brannon and Knoblauch (1982) discuss how, although teachers approach student writing with "the best intentions," the net effect of commentary is often to deny students authority over their own work. Keh (1990) refers to the high levels of frustration and time involved in creating written comments. In her seminal study, "Responding to Student Writing," Nancy Sommers (1982) notes that "paradoxically enough, that although commenting on student writing is the most widely used method for responding to student writing, it is the least understood. We do not know in any definitive way what constitutes thoughtful commentary or what effect, if any, our comments have on helping our students become more effective writers" (p. 148). Sommers' study of 35 teachers' comments finds that "…most teachers' comments are not text-specific and could be interchanged, rubber-stamped, from text to text" (p. 291), often lacking thoughtful engagement with the student and/or the subject about which they are writing. Chris Anson (1989) refers to responding to writing as "often difficult and tense" for both student and instructor in his introduction to *Writing and Response.* This edited collection encourages readers to consider response "a part of the social and interpersonal dynamics of the classroom community" (p.333); its contributors examine response from a variety of theoretical and practical contexts. Perhaps most succinctly summing up the problems with instructor commentary, White (1995), in *Assigning, Responding, Evaluating* states that "[f]ar too much of what teachers do with student writing is picky, arbitrary, unclear, or generally unhelpful" (p. 122), and that most faculty "repeat the patterns [of commentary] they have experienced as students" (p. 147).

Straub & Lunsford's (1995) *Twelve Readers Reading* presents a case study of twelve experienced instructors responding to a set of student texts and examines their comments in hopes that those artifacts from this group of "well-known, well-informed" instructors would provide models for other instructors hoping to improve their responses to student writing (p. 1). In 1996, following the publication of *Twelve Readers,* Straub extended his discussion of response in two articles. In "Teachers' Response as Conversation," Straub encourages instructors to focus more on ideas while consolidating formal concerns into only a few marginal comments. In "The Concept of Control," Straub examines the comments of several expert instructors as he encourages others to invoke the combination of directive and facilitative commentary that works best for them. The largest study pre-2000, Lunsford and Connors' 1993 "Teachers' Rhetorical Comments on Student Papers," gathered 26 readers to examine the "global comments"—those dealing with "issues of rhetoric, structure, general success, longitudinal writing development, mastery of conventional generic knowledge, and other large-scale issues" of 3,000 student papers (p. 206). Their conclusion—that there was much more work to be done to understand the range and context of commentary.

In the early 2000s, this wide-ranging and contradictory relationship with commentary continued. In 2006, Stern and Solomon sought to replicate and extend Connors and Lunsford's 1993 study referenced above. They found that not much had changed in instructor feedback since 1986 and called for more research in the efficacy and training of faculty to give feedback. Montgomery and Baker (2007) surveyed students, asked instructors to reflect, and examined the actual instructor commentary in light of those reflections, finding that perception of instructors didn't always match up with their actual commentary. Anson (2000) continues his pursuit of this topic when he explores the filters through which instructors respond to error in student texts and also encourages more reflective practices. Finally, Straub (2000) provides one of the most comprehensive overviews of commentary. Beginning with the responses from *Twelve Readers Reading,* he takes these responses, brings in others from newer instructors, and uses papers from one of his own composition courses as the subject for examining commentary in context. He also uses this material as the basis for creating a set of guidelines for responding to student writing.

While the body of texts produced in the 1980s and 1990s sought a greater understanding of commentary and, consequently, development of best practices, their effectiveness was hampered in several ways. First, they were, by and large, limited to small sample sizes. Other than Lunsford and Connors' 3,000 paper sample, most studies were confined to small, focused samples. Straub and Lunsford, one of the more comprehensive studies, dealt with 12 instructors, each of whom commented on 10 drafts. Second, as the workforce, especially in first-year writing, became increasingly populated by graduate students and NTT-faculty who more than likely did not have a degree in rhetoric, composition, or an allied field, it became more evident that research needed to occur to verify what was actually happening in real time, with real students and their instructors. Far too much of what was presented and published was still focusing on highly controlled experiment conditions and/or faculty with atypical experience in teaching writing. Consequently, then, all elements of the experiment, including the writing construct and the resulting feedback, were constrained. To begin to construct a more accurate illustration of the instructional process, we need to look at a program's current students, current instructors and use methodologies and technologies that will enable large-scale examination of the instruction that happens in a writing program.

Scholarship of the last two decades has started to fill this gap (Bailey & Garner 2010; Dixon & Moxley, 2013; Smith 1997), although the results aren't always as positive as some would like. Bailey and Garner's (2010) interviews of instructors find significant gaps between what instructors think commentary should be and what they find that it is. Summer Smith's 1997 examination of over 300 end comments "identified a complex set of commenting conventions, including a repertoire of primary genres and patterns of selecting from that repertoire to construct a product we recognize as an end comment" (p. 264). Smith cautions, however, in adhering too strictly to these conventions and notes that doing so may actually decrease the overall effect of the end comment because students will recognize the formula and disregard the comment as formulaic (p. 266). Cho et al. (2006) compared the work of peer reviewers to subject matter experts, but this work was more illustrative of peer comments since the subject-matter experts

were not course instructors, but others hired to provide commentary for the purposes of the study. Dixon and Moxley's 2013 study of nearly 120,000 rubric-based comments on approximately 18,000 papers demonstrated two distinctive trends: 1) instructors commented more on higher-level concerns than on sentence-level issues, and 2) given the weighting of format (i.e., documentation style) in the rubric, instructors made comments on format at a higher than expected level. Dixon and Moxley also commented at some length on the use of concordance and other text mining tools in writing research; they noted that while some might see them as developing a "Composition Panopticon, ... Deep Assessment in the context of our writing program represents a means to enable more reflexive, flexible, and beneficial communal practices. With digital tools and Deep Assessment, we can open the doors of the classrooms and offices to see which instructors exemplify best practices, and celebrate them, and which instructors are falling short, and encourage them" (p. 254).

Despite the research of the last 30+ years, it's clear that more can be done. Ferris (2014) surveyed and interviewed instructors at eight postsecondary institutions to see if they applied the concepts and practices suggested by prior research. Although Ferris found that most had, in fact, embraced these ideas, Ferris suggests that instructors should 1) pay more attention to what students do with feedback, 2) reflect on and evaluate their feedback practices, and 3) explore the impact of electronic feedback on their students' writing. The following year, Ferris launched the *Journal of Response to Writing*; Ferris explains that the nearly decade-long gap in research in response, except in L2 and linguistics journals, prompted this launch. Ironically, in the ensuing three years, a number of significant studies (Anson & Anson, 2017; Anson & Moore, 2016; Cohn & Stewart, 2016; Johnson, Wilson, & Roscoe, 2017; Laflen & Smith, 2017; Lancaster, 2016; McGrath & Atkinson-Leadbeater, 2016; Ruggiero, 2017; Simpson, 2017; Yoon et al., 2016) have continued to examine the role and practice of instructor feedback from a variety of perspectives.

Some of the above studies have also included significant corpuses of data in doing so. However, much work remains to create meaningful intersections of corpus analysis and writing studies. Some of the most recent work in this direction includes Aull and Lancaster (2014) and Aull (2015), who engage corpus analysis to compare first-year writers' work to more experienced and expert writers. Both of the aforementioned studies find important differences in the work of the new and expert writers; Aull's 2015 book concludes by discussing pedagogical implications of corpus analysis. Aull encourages a pedagogy in which corpus-based findings are shared and analyzed with students to help them understand conventions of academic writing and their relation to those. She notes that:

> At its best, corpus-based study of FY writing supports the right of all developing academic writers to a fair share of academic writing and knowledge-making, based on the idea that transparency about what is privileged therein is one form of access thereto. (162)

In another instance, Staples et al. (2016) examined phrasal as well as clausal features to understand the development of writers in Britain from first-year to graduate students, finding

evolution in use of both features as students progressed in their university education, but that this evolution differed by discipline and genre. Moxley and Eubanks (2015) examined feedback given by both peers and instructors; the feedback, though, focused primarily on rubric scores but represents the largest of this type of study to date. And in one of the most recent studies on instructor feedback, Adel (2017) examined instances of metadiscourse in such feedback and found that discourse as "problem/solution-focused," rather than serving the more typical discourse-organization focus. The field needs to build on these recent efforts and continue to conduct large-scale studies of instructor feedback that examine the work of actual instructors rather than experts modeling best practices. It also should continue to look at electronic feedback, since an increasing number of courses and institutions have integrated learning management systems (LMS's) or other tools into their daily operations.

This project builds upon prior research in instructor commentary by examining a large corpus of instructor feedback from a single writing program using the lexicon developed via the research of Anson and Anson (2017). Anson and Anson (2017) sought to examine both peer and instructor response to student writing by 1) surveying experienced writing instructors nationally to develop a lexicon of terms included in effective instructor responses, and 2) analyzing a corpus of instructor and peer responses from a single writing program to see to what extent each group of responders incorporate these key terms in their responses to student writing. This study takes Anson and Anson's lexicon and replicates their work by examining a body of work of new instructors against the lexicon to determine whether new instructors' feedback more closely resembles the lexicon of novice respondents or that of more experienced writing instructors. Further, it studies four cohorts of new instructors for trends over time. Finally, it tracks evolution in the cohorts' vocabulary over their two years in the program. This information may prove valuable to others in a similar position of directing large writing programs.

3.0 Research Questions

This study applies the Anson and Anson lexicon to corpus data from instructors in a first-year writing program over five years (10 semesters) from August 2012 through May 2017. The university is a public, R1 (Doctoral University—Highest Research Activity) institution located in the Southwest. The first-year writing program is a two-semester sequence, delivered as a hybrid. Students met with their classroom instructor once weekly for 80 minutes in a lecture/workshop environment. During each course, they completed a series of scaffolded assignments that were evaluated by the members of their grading groups—generally four to six instructors, who were either MA or PhD students, or non-tenure track instructors. The courses focused on what we considered key genres of academic writing—rhetorical analysis, literature review, and researched argument. Students turned in all assignments, except weekly homework assigned by their classroom instructor, through Raider Writer, our online learning management system. Students accessed their graded assignments in Raider Writer as well.

The first course, ENGL 1301, is designed to help build the foundation for each student's academic writing career. Throughout the semester, students complete a variety of assignments in

order to practice and improve their critical thinking, writing, and reading skills. Students write summaries of texts and visuals, evaluations of academic sources, and a rhetorical analysis of a particular text. In doing so, they learn more about planning and organizing writing, drafting and revising to improve content and coherence, and editing and proofreading to increase their understanding of conventions of grammar and mechanics. Finally, they develop a greater understanding of the role of writing in various university courses.

The second course, ENGL 1302, focuses on persuasive writing and writing from sources. Students build on the skills learned in English 1301 as they conduct preliminary research and write a literature review, develop claims for argument, compile and evaluate evidence and support for their claims, learn to recognize and avoid fallacious reasoning, and gain a better understanding of the role of language in argument. Students also conduct academic research using both print and electronic sources, evaluate and incorporate source material into an argument, and practice citing that material appropriately.

A key feature of the courses is the incorporation of the grading groups. In order to enable students to receive more commentary on their writing without overwhelming the classroom instructor, the program uses small groups of instructors, some of whose primary assignment is to read and respond to student writing. Shorter assignments, including peer critiques, and writing reviews which make up 75 – 80% of the writing, are graded by one instructor, and all major drafts in each course (2 in 1301 and 4 in 1302) receive at least two reads in order to obtain an averaged score. The distribution, evaluating, and archiving of student assignments is facilitated by our LMS.

Given that a significant number of instructors were new to the program each year, it was important to integrate them into the system as quickly as possible. Knowing whether and how the new cohort integrated was generally ascertained by a variety of indirect measures, primarily by talking with new and more experienced cohorts in workshops, and by reviewing samples of each instructor's comments. Using the two lexicons developed via Anson and Anson's survey provides another way by which the new instructors' work may be understood in a larger context.

To that end, this study asked the following questions:

1. Does the work of new instructors (MA1s) more closely resemble the lexicon of novice or experienced responders to student writing?
2. How does the new instructors' work compare to that of more experienced (PHD1 or INS) instructors in the program throughout their time?
3. How does their work evolve over a four-semester longitudinal time frame (as MA1 or MA2 experience levels) in the first-year writing program?

4.0 Research Methodology

4.1 Corpus

The corpus of assignments and instructor commentary was collected in the regular course of instruction in the program via the internally constructed program management system. All

enrolled students in ENG1301 and ENG1302 submit assignments in the application and instructors view, comment on, and grade assignments within the same web-based application. All assignments, comments, and other associated data are stored and later archived in an SQL database. Over the ten semesters included in this study, approximately 12,000 students were enrolled in the two courses. Both courses in the program used a common syllabus, readings, and writing assignments. Each semester, all instructors of both courses, regardless of years of experience, attended several professional development sessions. These sessions included conceptual training in abstract rhetorical concepts; they also examined and discussed assignment descriptions, learning objectives, and response criteria. Instructors also worked with sample assignments and feedback during the professional development workshops.

Each course required between 13 and 16 graded assignments. Most assignments were between 300 and 600 words in length; two assignments in the first semester course and four assignments in the second semester course were between 1,200 and 2,000 words in length. For the purposes of this study, two assignments from each course were included in this corpus; these assignments represented the first and last long assignments (1,200 – 2,000 word) in each course to which instructors responded.

4.2 Population

The overall university population currently exceeds 36,000 and has been increasing steadily for over a decade. Students in this program represent on average approximately 33% to 36% of the entering first-year class, which has averaged between 6,300 and 6,800 during the five years (2012 – 2017) of this study. Over this same time frame, the population of White students decreased from 62% to 57%, and the Hispanic student population increased from 21% to 25%. Nearly 85% of students were from the home state of the institution, and approximately half of each entering class came from the top 25th or better percentile of their graduating class.

Instructors in this program also followed the typical demographic profile of graduate students at this institution (a majority of whom are White students, with international students and Hispanic students making up the next largest groups). These students were either graduate students at the master's or doctoral level, or they were non-tenure track faculty. Most of the faculty had master's degrees; a few, however, did hold doctorates. Nearly all of the incoming master's students taught in the program for the duration of their studies. Doctoral students taught in the program during their first year before moving on to sophomore-level courses in literature, creative writing, or technical communication. Some did return to teach in the first-year writing program later in their time at the university, but it was never required that they do so. The non-tenure track faculty's experience ranged from three years of teaching to over 50 years. A core of about a dozen had worked with the program for over 15 years at the time this study began.

Instructors included in the study are classified in one of four groups: first-year MA/MATC students (MA1), second year MA/MATC students (MA2), first-year doctoral students (PHD1), or non-tenure-track teaching faculty (INS). Second-year doctoral students occasionally teach in the program, but they did not comprise enough of the population to warrant inclusion, as they

commented on less than 3% of the submitted corpus. Table 1 shows the breakdown of the instructor population.

Table 1

Instructor Population

Semester	Instructors (3+ years' experience)	First-year MA/MATC	Second-year MA/MATC	First-year PhD students	Second-year PhD students
FA12	18	16	10	8	3
SP13	20	11	10	7	2
FA13	25	14	9	13	3
SP14	22	12	8	13	2
FA14	36	19	12	15	3
SP15	26	16	13	13	1
FA15	33	16	14	8	5
SP16	35	13	17	8	5
FA16	38	10	16	7	3
SP17	35	10	16	6	3

The assignments included in the study sample were two drafts of the rhetorical analysis from ENGL 1301 and the first (the literature review) and final (the researched argument) major drafts from ENGL 1302. These assignments ranged in length from approximately 1,200 words (the initial draft of the literature review) to 2,000 words (the final draft of the researched argument). I chose these four assignments because they represent the most typical challenge for new instructors— that is, commenting on multiple aspects of an extended work. For each assignment, instructors provide formal commentary on a preliminary and a final draft. Rather than targeting a particular segment of the process, such as identifying rhetorical choices made by an author or creating an annotated bibliography of potential sources, these assignments demand that students bring everything that they've been working on in the course together in one place. Our instructors are tasked with reading and prioritizing concerns to discuss with students in each of these assignments. The breakdown of assignments graded, discrete comments made, and unique and total words generated by each instructor group for the sample assignments is shown in Table 2. The sample for the study consisted of 17,534 pieces of student writing representing 141,659 discrete comments on that writing, with 58,300 unique words out of over 8.25 million total words written.

Table 2

Breakdown of Assignments Graded, Comments, and Words by Population

	Instructors (3+ years' experience)	First-year MA/MATC	Second-year MA/MATC	First-year PhD students
Assignments	8,320	4,148	2,576	1,947
Discrete comments	64,326	33,388	21,525	17,859
Average comment/document	7.7	8.0	8.3	9.2
Unique words	44,102	17,855	13,710	16,954
Total words	3,979,300	1,936,851	1,183,919	931,763
Average words/comment	61	58	55	52
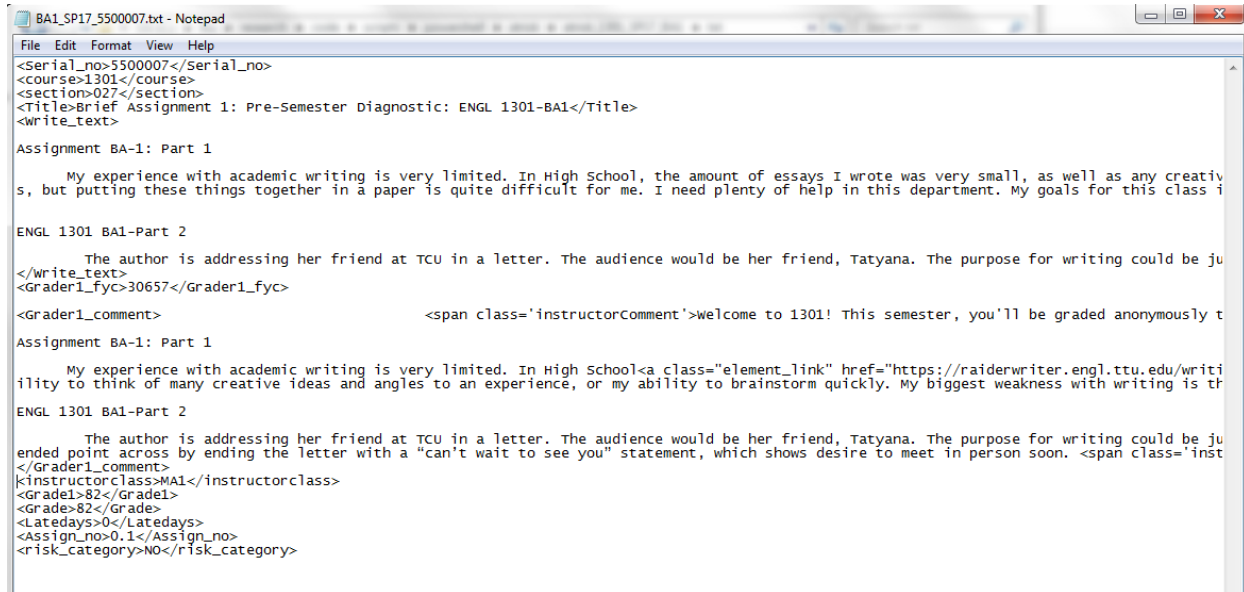
4.3 Process

The study was deemed exempt by the IRB board because the data was deidentified, collected during the typical course of instruction in these classes, and deemed to pose no risk to the writers if used for this research. The LMS used an SQL database to store all student writing, instructor comments, and associated information. To begin the study, the instructor comments along with other associated information, such as assignment identifier and instructor level (e.g., MA1, MA2, INS) were extracted from the program archives in the SQL database using PowerShell scripts. Readers will note that far more fields than used in this particular study can be extracted. This script shown in Figure 1 extracts the following fields from the database, once the parameter of semester, course, and assignment number are set.

```
•     Serial no [of submission]
•     Course
•     Section
•     Title [of assignment]
•     Write text [student submitted text in response to assignment]
•     Grader1_fyc [ID number assigned to grader]
•     Grader1_comment [grader's comments on the assignment]
•     Instructorclass [MA1, MA2, PHD1, PHD2, or INS]
•     Grade1 [grade assigned by the instructor]
•     Grade [final grade, reflects any adjustments by classroom instructor
or late penalties]
•     Latedays [number of days late—'0' if on time]
•     Assign no [identifier of which assignment this is in the course]
•     Risk category [students are classified each term as new, at risk
(GPA below 2.5) or in good standing (GPA above 2.5, no academic suspension or
probation)
```

Figure 1. Fields extracted for each writing assignment.

This information from every assignment in the set (i.e., all assignment #2s from ENGL 1301 from Fall 2016) was extracted into a separate .txt file and stored in a single folder. A sample extracted file is shown in Figure 2.

Figure 2. Plain text example of data extracted.

Note that each field extracted from the SQL was enclosed by HTML tags. These tags were used in the import process to the text mining software, Provalis software's ProSuite, to designate variables.

Provalis Software's ProSuite includes QDA Miner (qualitative data analysis), SimStat (statistical analysis and bootstrapping), and WordStat (content analysis and text mining) software. The three applications work from a single data import, so the data for this project was imported via QDA's Document Conversion Wizard. Following the import into QDA Miner (4.1.37; 2014), data were analyzed using QDA Miner and Wordstat (7.1.19; 2014). Data were stemmed (using the same Porter stemmer as Anson & Anson) and examined first as a full corpus—all 17,000 documents over 10 semesters with all instructor groups. Data was then filtered by:

- Semester (e.g., FA12, SP13)
- Assignment (these had identifying assignment numbers—Draft 1.1 and 1.2 (first semester) and 1.1 and 2.2 (second semester)
- Instructor experience level (MA1, MA2, PHD1, and INS)

Each set of comments was then run against no dictionary except standard stopwords (note—the pronoun "you" was included as I wanted to see if there was a difference in how instructors addressed authors, directly or via the more generic reference to the text [paper, essay, draft, etc.]). Next, the comment sets were run against Anson and Anson's "novice" and "principled"

lexicons. The terms "novice" and "principled" arise from these questions posed by Anson and Anson: "Does peer response focus on the same writing-related concepts as the response students have received from their instructors, representing the transfer of learning from instructor to student to peer? Or does peer response contain little substantive content, or content unrelated to the principles emphasized by experts in the discipline of writing studies?"

In the survey they conducted to generate these lists, Anson and Anson again refer to these terms. In the survey, they invite participants (all of whom are or have an interest in writing program administration with experience in teaching college-level writing) to "'think about concepts that are important for good, principled response to writing,' as well as concepts 'that might be likely to appear in beginners' response to writing'" (p.14). The top responses to the above appear in Table 3.

Table 3

Novice and Principled Lexicons

Novice lexicon (stemmed)	Principled lexicon (stemmed)
AWKWARD	ARGUMENT
COMMA	AUDIENC
CORRECT	CLEAR
FLOW	COHER
GOOD	DEVELOP
GRAMMAR	EVID
PUNCTUAT	FOCUS
SENTENC	IDEA
SPELL	ORGAN
UNCLEAR	READER
	SPECIF
	STRUCTUR
	SUPPORT
	THESI

The resulting matrices that represented the MA1 students were examined and analyzed to establish entry points for each cohort. Each cohort was then examined in terms of 1) comparison to each other, 2) evolution over its two years in the program, and 3) comparison to the experienced INS of their cohort years. Descriptive frequency analyses, *t*-test comparisons for lexicon lists and individual terms, and correlational analyses were conducted.

5.0 Results

5.1 RQ1 Results

RQ1: Does the work of new instructors (MA1s) more closely resemble the lexicon of novice or experienced responders to student writing?

The work of the new instructors (MA1s) incorporates more of the terms associated with principled response than novice response. In terms of raw percentage of novice terms used, MA1s across the four cohorts used three of the ten terms in more than 1% of their comments and used two of the ten terms, good and sentence, more than 10% of the time (19.9% and 16.3%, respectively). The use of the other five terms ranged from 0.43% to 0.83%. Of the four cohorts of MA1s included in the study, proportionately higher use of terms such as awkward, comma, correct, and spell occurred during the first three cohort years from 2012 – 2015.

In terms of raw percentage of principled terms used, MA1s across the four cohorts used one term (coher) less than 0.50%, four terms between 2.5% and 5%, six terms between 5% and 10%, and three terms in more than 12% of their comments. See Tables A1 and A2 in the Appendix for full data.

5.2 RQ2 Results

RQ2: How does their work evolve over their four semesters (as MA1 or MA2 experience levels) in the first-year writing program?

5.2.1 Novice terms.

- Overall, of the ten novice terms, use of six of the ten (*awkward*, *comma*, *correct*, *punctuat*, *spell*, and *unclear*) decreased from the cohorts' first to second year, two increased (*good* and *grammar*), and two (*flow* and *sentence*) remained virtually unchanged.
- Comparing cohorts: In looking at the number of times and number of novice terms used more than 0.10% above the mean, the use of the novice terms by the cohorts declined slightly in two cohorts (from 14 to 12 and 16 to 13), increased slightly in one cohort (from 9 to 12), and increased sharply (from 10 to 21) in one cohort.
- Comparing cohorts: In looking at above average use of the terms at each of the four drafts, two of the cohorts (2012 – 2014 and 2014 – 2016) evidenced more than a two-term increase or decrease in above average use. The 2012 – 2014 cohort increased from two to five terms used above average on draft 1.1 in 1301, but decreased from seven to two terms used above average on draft 1.2. The 2014 – 2016 cohort increased from two to six terms on 1301's 1.2 draft, and from three to six terms on 1302's 1.1 draft. See Tables A3 and A4 in the Appendix for full data.

5.2.2 Principled terms. Overall, of the 14 principled terms, use of four of them (*audienc*, *clear*, *develop*, and *specif*) increased from the cohorts' first to second year and eight (*argument*, *evid*, *idea*, *organ*, *reader*, *structure*, *support*, and *thesi*) decreased; two *coher* and *focus*) remained virtually unchanged.

- Comparing cohorts: In looking at the number of times and number of principled terms used more than 0.10% above the mean, use of the principled terms increased in all four of the cohorts (from 7 to 16; 7 to 19; 8 to 12; and 10 to 20, respectively).
- Comparing cohorts: In looking at above average use of the terms at each of the four drafts, three of the cohorts (2012 – 2014, 2013 – 2015, and 2015 – 2017) evidenced more than a two-term increase or decrease in above average use. The 2012 – 2014 cohort increased from zero to four terms used above average on draft 1.1 in 1302 and increased from three to seven terms used above average on 1302's draft 2.2. The 2013 – 2015 cohort increased use from one to five terms on 1301's draft 1.1, one to four terms on 1301's draft 1.2, and three to six terms on 1302's draft 2.2. The 2014 – 2016 cohort showed the least number of increases, only moving from one to four terms on 1301's draft 1.1. The final cohort, 2015 – 2017, increased from two to seven terms on 1301's 1.1 draft, and from zero to five terms on 1301's 1.2 draft.

5.3 RQ3 Results

RQ3: How does their work compare to that of more experienced (PHD1 or INS) instructors in the program throughout their time?

5.3.1 Descriptive Frequency: Comparison of all classifications using novice and principled lexicons. Based on Figure 3 and Figure 4 below, it is clear that the MA1s, as well as the rest of the instructor groups, make greater use of the Ansons' "principled terms" lexicon. The first two figures below detail the averages over the five years. Only two of the novice terms, *good* and *sentence*, are used by any of the instructor experience cohorts in more than 5% of their comments.
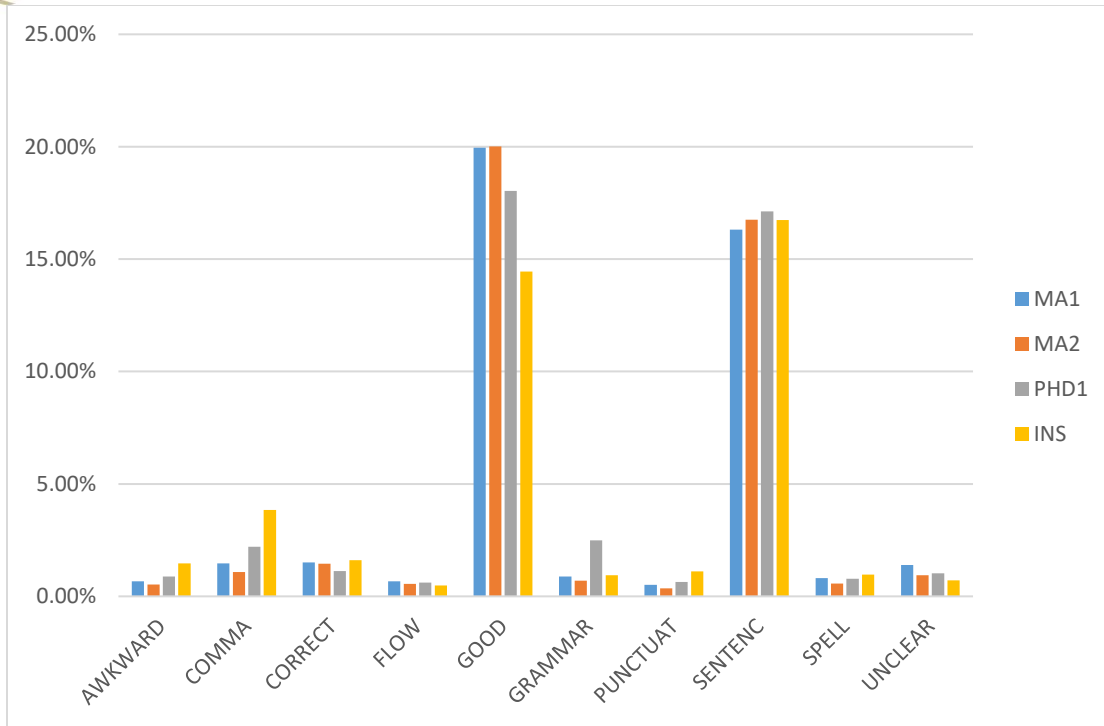
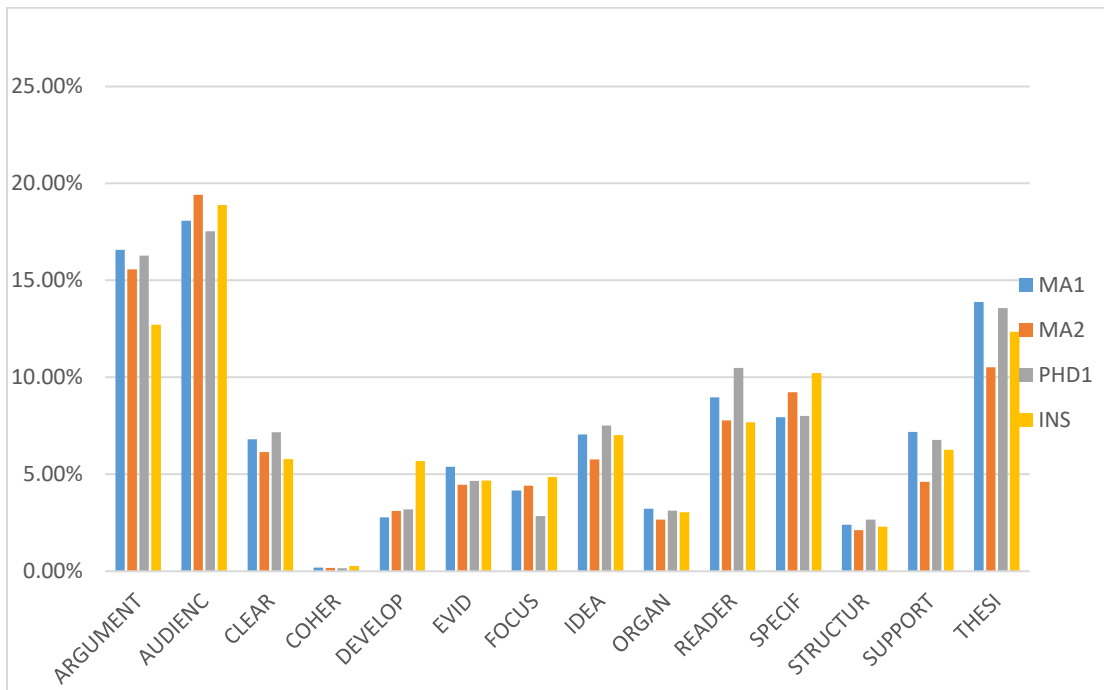*Figure 3.* Overall average use: Novice lexicon.

Figure 4. Overall use: Principled lexicon.

5.3.2 T-test results. MA1 to INS (across all cohorts):

The highlighted terms in Table 4, below, showed significance between the MA1 cohorts and Instructor cohorts. Those showing significance were as follows: The novice terms *awkward* ($p < 0.001$), *good* (0.001), and *unclear* (0.05) showed significance in two-tailed *t*-test. The principled terms *develop* (0.001) and *specif* (0.05) also showed significance.

Table 4

P-values for First-year MA Students and Instructor Cohorts

Word	% MA1 comments	% Instructor comments	*P*-value
AWKWARD	0.01	0.02	0.00
COMMA	0.02	0.04	0.00
CORRECT	0.02	0.02	0.67
FLOW	0.01	0.01	0.24
GOOD	0.19	0.14	0.00
GRAMMAR	0.01	0.00	0.89
PUNCTUAT	0.00	0.01	0.00
SENTENC	0.16	0.16	0.71
SPELL	0.01	0.01	0.26
UNCLEAR	0.02	0.01	0.03
ARGUMENT	0.16	0.12	0.37
AUDIENC	0.18	0.17	0.88
CLEAR	0.06	0.05	0.27
COHER	0.00	0.00	0.42
DEVELOP	0.03	0.06	0.00
EVID	0.06	0.04	0.47
FOCUS	0.04	0.05	0.31
IDEA	0.07	0.08	0.97
ORGAN	0.03	0.03	0.73
READER	0.09	0.08	0.29
SPECIF	0.08	0.09	0.02
STRUCTUR	0.03	0.02	0.79
SUPPORT	0.07	0.06	0.60
THESI	0.13	0.12	0.44

5.3.3 Correlational analyses. For use of the novice lexicon, correlation was very strong between MA1s and INS in all cohorts (0.98). A similar result was found for the expert lexicon (0.96). Between the MA1 and MA2 years in all cohorts, the correlation of novice terms was 1.0 and for expert terms, 0.97. Correlations are summarized in Table 5.

Table 5

Correlation Summary

	MA1/INS	MA1/MA2
Novice	0.975	0.999
Principled	0.96	0.968

6.0 Discussion

In 1982, Nancy Sommers, in summarizing recently completed research on styles of commenting, found that "the news from the classroom is not good" (p. 154). Her research found that instructors often focused on grammar, responded negatively to student writing, and didn't focus enough in those early assignments on providing genuine guidance for revision—and these were largely experienced instructors. In their defense, she notes that "they [the teachers] told us that responding to student writing was rarely stressed in their teacher-training or in writing workshops" (p. 154). Those teachers were simply applying to student writing what they had been taught in terms of reading and interpreting literary texts.

Fast forward 36 years—what differences, if any, does sustained professional development have, even on far less experienced instructors? And what do contemporary methods of data storage, retrieval, and analytics help us learn about those instructors' evolution?

Overall, the results reveal a number of findings about the feedback of both the most novice and more experienced instructors in this program.

1. While there are some differences in commentary as seen via examination of the two lexicons, the differences are perhaps less than one might assume.
2. The cohorts do increase their use of the principled terms as they move through the two years' appointment in the program, but few of the increases demonstrate statistical significance.
3. Few of the terms from either the novice or principled lexicon, with the exception of terms that also appear in the assignment descriptions, what I label as "content terms," appear frequently in the overall corpus. Figure 6.1 below indicates the only terms used in more than 10% of the comments of the entire corpus across all levels.

As Figure 5 illustrates, *audience*, *good*, *sentence*, *thesis*, and *argument* are the only terms from either lexicon used in more than 10% of the comments overall. In Figure 6, the most-used

terms from the most experienced instructors, we again see the same five terms from the two lexicons: *audience*, *thesis*, *good*, *sentence*, and *argument*. And for the new instructors, Figure 7 reveals the same terms from the principled (*audience*, *argument*, and *thesis*) and novice (*good* and *sentence*) as from the experienced instructors.
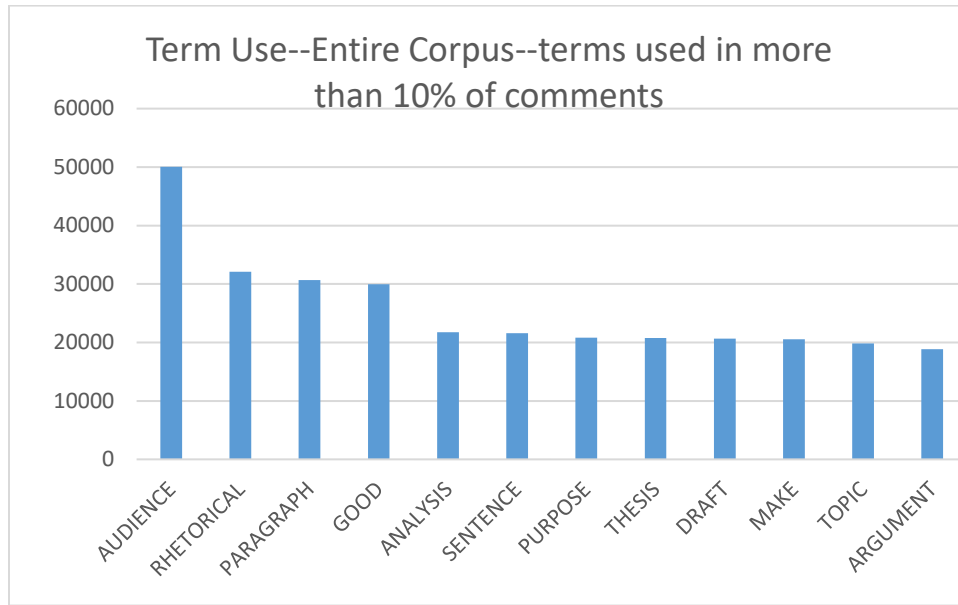

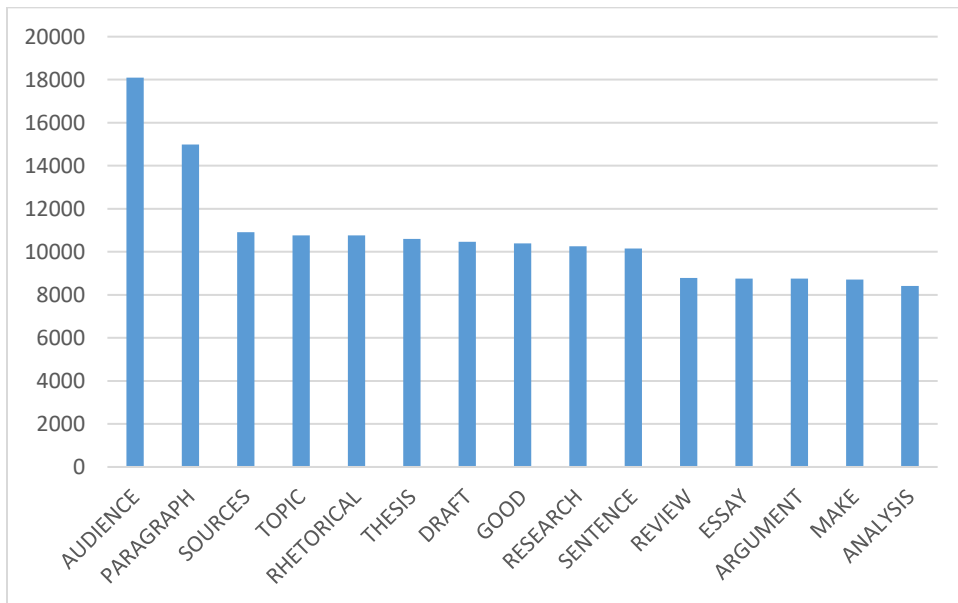*Figure 5.* Entire corpus: Term use in more than 10% of comments.

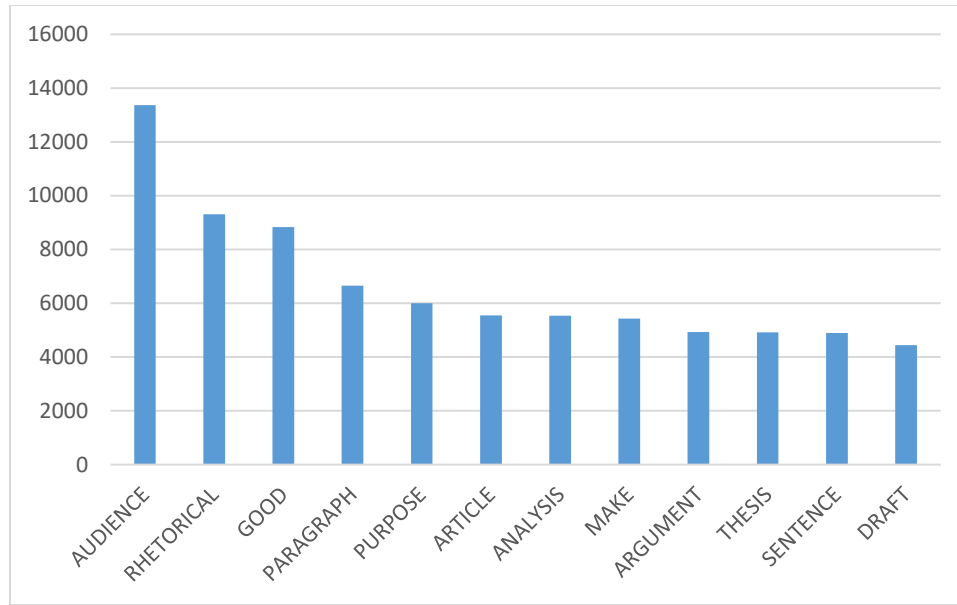*Figure 6.* INS Full Corpus: Terms from 18% to ~10%.

Figure 7. MA1 corpus: All semesters, full list 24.7% - 8.2

Based on the above metrics, it would appear that instructors in this program had acquired a more consistent vocabulary, but not primarily one based on Anson and Anson's two lexicons—instead, the most frequent and commonly used terms seem to come from a more local "canon," that is, one based on the assignment descriptions and course outcomes. Regardless of whether the acquisition of a common vocabulary came from more global concepts or an assignment-based local canon, using common terms is something that Nancy Sommers (1982) saw as contributing to "thoughtful commentary" on student writing. The fact that the majority of the terms in either lexicon are used in no more than 5 – 10% of instructor comments may cause us to question how "common" they actually are or how much weight they hold compared to a local canon. One thing to consider is that all four instructor groups averaged between 7.7 and 9.2 discrete comments per document. It is possible that these terms did appear in the vast majority of assignments that were evaluated, but in only one of the comments provided.

It is also possible that researchers need to look more closely at the connections between what I'll term "writing" lexicons and "content" lexicons—words connected closely to whatever the assignment is being evaluated for—in order to see how those lexicons interact. As one brief example, the two most-frequently used terms in the principled lexicon overlap with content lexicons for two major assignments in the FYC sequence. For example, the last of the four drafts whose comments are included here is a researched argument. While overall, MA1s used the term "argument" in nearly 16% of their comments, that number dropped to just over 9% on the remaining three drafts. Similarly, the MA1s used the term "audience" in nearly 18% of their total comments but in just 2% of their ENGL 1302 comments. In ENGL 1301, "audience" was a prominent part of the rhetorical analysis assignment, and use averaged nearly 34% in those comments on both the preliminary and final draft.

As no one has previously studied how quickly new instructors acquire a professional vocabulary for responding to student writing, it is hard to know whether or not the results of this particular group of instructors would be considered "typical." However, it may well be that the context of this writing program contributed to a more accelerated acquisition. New instructors began their professional development activities a month or so before arriving on campus, by completing selected readings, as well as drafting and submitting several of the first semester writing assignments to the program's director. Feedback from incoming instructors each year indicated that many of them had not completed these types of assignments in any of their own prior coursework before, and thus these would have been teaching assignments they had never experienced. The common syllabi, reading selections, and writing assignments provide all instructors with a nearly identical entry point to the courses. Additionally, the collaborative nature of the hybrid courses and the grading groups, who met virtually or face-to-face to discuss assignments prior to evaluating them, may have expedited the MA1s use of a common vocabulary. The web application developed for program use also contained several features that allowed real-time chat between instructors while grading and enabled all members of each grading group to examine recent comments written by their group members on assignments. Finally, all instructors had access to a comprehensive support website for the courses; this site offered lesson plans, detailed instructions for evaluating each assignment, and other resources for instructors to access 24/7. It is plausible to think that removal of one or more of these elements would lead to more and significant differences in the new instructors' (and program instructors in general) use of the two lexicons.

6.1 Limitations of the Study

As with any research, limitations exist. In terms of population and corpus, this study only looks at one American institution's writing program, though it does so over five years. Additionally, it did not look at, literally, the "first" assignment that the instructors responded to (the "first" was really about 6 – 7 weeks in). This second limitation *could* be remedied by extracting the first assignment's comments for the ten semesters, though the first assignment of the semester differs significantly in scope and direction from the others included in the study. And while such options as designing a pre-test to assess the MA1 instructors' initial vocabulary prior to any training might provide more useful information, finding the time to do this during their initial days on campus may be difficult, given the myriad of demands on their time.

7.0 Conclusions

In the last several years, researchers in writing studies have returned to and advocated more strongly for the development of a more common vocabulary when talking about and teaching writing (most comprehensively Adler-Kassner & Wardle, 2015). Program administrators have likewise pointed out the need for ongoing models of TA training that break "the accepted pattern of the one and done training" (Obermark, Brewer, & Halasek, 2015, p.34). The program

examined here, as part of the mandatory professional development and programmatic structure, including the web application and the collaborative grading groups, likely contributed to the apparently rapid assimilation of the new instructors. Although not the focus of this study, the ongoing requirement that all instructors participate in professional development and the embedded, tacit mentoring that the grading groups provide also likely played a role in the similarity in term use from both lexicons across the board.

The study also responds to larger calls in the literature of assessment, including Earl and Timperly for "understanding how educators at all levels actually use evidence in their thinking and their decision-making" (p.1). Earl and Timperly argue for evidence-based conversations which are, in their view, iterative processes of "asking questions, examining evidence, and thinking about what the evidence means in the particular context" (p.3). Looking carefully at the local context of the evidence, in this case the comments made by instructors, mandates that administrators and instructors embrace the opportunity to view the activity of grading and commenting on student work as something else—formative assessment. Bennett (2010) discusses formative assessment as a yet-to-be understood concept. Ultimately, he considers formative assessment to be a coherent (and I would add, complex) system that works as assessment for learning—not as assessment of learning. Given the history of the roles of formative and summative commentary in writing pedagogy, I'd suggest that program administrators explore the formative in all of its potential contexts.

8.0 Directions for Future Research

Working with the lexicons developed via Anson and Anson's survey is a useful starting point for understanding more of what our instructors actually do when responding to student writing, as well as for identifying critical differences in our instructors' comments. The lexicons, though, only provide us with a small subset of expected (thus acceptable) terms included in commentary—terms that afford students the opportunity to act upon receiving them via revision or transfer. More research is necessary, though, to expand and refine the lexicons and their impact on student writing. One possibility is to return to the current data set to engage in additional lexical analysis of both the novice and principled lexicons as well as the overall frequency tables to understand how terms are used in the context of response by the various instructor groups. Differences in the application of the terms might help us understand why comments might be labeled as more or less helpful to writers. For example, differentiating between terms such as *clear* and *unclear* and their use in phrases by instructors would lead to a more nuanced analysis. As well, a more detailed, focused analysis of lexical and phrasal overlap between the MA1 and INS groups would be useful. Another option is to examine the data in terms of markers of stance; finally, topic modeling could be used to locate more subtle differences in the instructor comments that aren't as easily identifiable with lexical analysis. Such examinations could serve as a baseline for broadening the study out to other sets of assignments and commentary, perhaps helping us build a set of threshold concepts for talking about writing with our students. Finally, employing other methodologies such as case studies and

ethnographies of individual instructors at various experience levels would allow us to follow the evolution of particular individuals' instruction and perhaps generate hypotheses that could then be examined in the context of larger data sets such as the one used in this study.

It also makes sense to replicate and expand Anson and Anson's survey to other stakeholder groups. As with much research on the teaching of writing, we default to the group most accessible to us—other writing professionals. Replicating this survey with other stakeholders— graduate teaching assistants and undergraduate students at both lower and upper division levels—could help us understand whether or not a gap exists in understanding what constitutes good feedback from the various stakeholders. Additionally, expanding the population of the study to include writing and feedback from other disciplines and institutions from across the United States and international writing programs could help us create a more comprehensive portrait of instructional practices.

Author Biography

Susan Lang currently serves as the Director of the Center for the Study and Teaching of Writing at The Ohio State University. For the last 25 years, her research has focused on writing program administration, especially instructor training and assessing student writing on a large scale, and on online teaching and learning, including software development. Her work is available in journals addressing composition studies, writing program administration, and technical communication. She is the 2016 Kenneth Bruffee Award winner for "Taming Big Data through Agile Approaches to Instructor Training and Assessment: Managing Ongoing Professional Development in Large First-Year Writing Programs."

References

Ädel, A. (2017). Remember that your reader cannot read your mind: Problem/solution-oriented metadiscourse in teacher feedback on student writing. *English for Specific Purposes*, *45*, 54–68.

Adler-Kassner, L., & Wardle, E. (2015). *Naming what we know: Threshold concepts of writing studies*. Boulder, CO: University Press of Colorado.

Anson, C. M. (1989). *Writing and response: Theory, practice, and research*. Urbana, IL: National Council of Teachers of English.

Anson, C. M. (2000). Response and the social construction of error. *Assessing Writing*, *7*, 5–21.

Anson, C. M., & Moore, J. (2016). *Critical transitions: Writing and the question of transfer*. Anderson, SC: Parlor Press.

Anson, I. G., & Anson, C. M. (2017). Assessing peer and instructor response to writing: A corpus analysis from an expert survey. *Assessing Writing*, *33*, 12–24.

Aull, L. (2015). *First-year university writing: A corpus-based study with implications for pedagogy*. New York: Springer.

Aull, L. L., & Lancaster, Z. (2014). Linguistic markers of stance in early and advanced academic writing: A corpus-based comparison. *Written Communication*, *31*(2), 151–183.

Bailey, R., & Garner, M. (2010). Is the feedback in higher education assessment worth the paper it is written on? Teachers' reflections on their practices. *Teaching in Higher Education, 15*(2), 187–198.

Brannon, L., & Knoblauch, C. H. (1982). On students' rights to their own texts: A model of teacher response. *College Composition and Communication*, *33*(2), 157–166.

Cho, K., Schunn, C. D., & Charney, D. (2006). Commenting on writing typology and perceived helpfulness of comments from novice peer reviewers and subject matter experts. *Written Communication*, *23*(3), 260–294.

Cohn, J. D., & Stewart, M. (2016). Promoting metacognitive thought through response to low-stakes writing. *Journal of Response to Writing*, *2*(1), 58 – 74.

Connors, R. J., & Lunsford, A. A. (1993). Teachers' rhetorical comments on student papers. *College Composition and Communication*, *44*(2), 200–223.

Dixon, Z., & Moxley, J. (2013). Everything is illuminated: What big data can tell us about teacher commentary. *Assessing Writing*, *18*(4), 241–256.

Ferris, D. (2015). A catalytic event for response research? Introducing our new journal: Editor's introduction. *Journal of Response to Writing*, *1*(1), 1–9.

Ferris, D. R. (2014). Responding to student writing: Teachers' philosophies and practices. *Assessing Writing*, *19*, 6–23.

Johnson, A. C., Wilson, J., & Roscoe, R. D. (2017). College student perceptions of writing errors, text quality, and author characteristics. *Assessing Writing*, *34*, 72–87.

Keh, C. L. (1990). Feedback in the writing process: A model and methods for implementation. *ELT Journal, 44*(4), 294–304.

Laflen, A., & Smith, M. (2017). Responding to student writing online: Tracking student interactions with instructor feedback in a Learning Management System. *Assessing Writing*, *31*, 39–52.

Lancaster, A. (2016, October). Responding to writing through instructor screencasts: Cognitive walkthrough, reader response, and student-centered access. In *Professional Communication Conference (IPCC), 2016 IEEE International* (pp. 1-5). IEEE.

Lang, S. (2016). Taming big data through agile approaches to instructor training and assessment: Managing ongoing professional development in large first-year writing programs. *Writing Program Administration: Journal of the Council of Writing Program Administrators, 39*(1), 81–104.

Lang, S., & Baehr, C. (2012). Data mining: A hybrid methodology for complex and dynamic research. *College Composition and Communication, 64*(1), 172–194.

McGrath, A., & Atkinson-Leadbeater, K. (2016). Instructor comments on student writing: Learner response to electronic written feedback. *Transformative Dialogues: Teaching & Learning Journal*, *8*(3), 1 - 16

Montgomery, J. L., & Baker, W. (2007). Teacher-written feedback: Student perceptions, teacher self-assessment, and actual teacher performance. *Journal of Second Language Writing, 16*(2), 82–99.

Moss, P., Pullin, D., Gee, J., Haertel, E., & Young, L. (Eds.). (2008). *Assessment, Equity, and Opportunity to Learn* (Learning in Doing: Social, Cognitive and Computational Perspectives). Cambridge: Cambridge University Press.

Moss, P., Girard, B., & Greeno, J. (2008). Sociocultural Implications for Assessment II. In P. Moss, D. Pullin, J. Gee, E. Haertel, & L. Young (Eds.), *Assessment, Equity, and Opportunity to Learn* (Learning in Doing: Social, Cognitive and Computational Perspectives, pp. 295-332). Cambridge: Cambridge University Press.

Moxley, J. M., & Eubanks, D. (2015). On keeping score: Instructors' vs. students' rubric ratings of 46,689 essays. *WPA: Writing Program Administration*, *39*, 53–80.

Obermark, L., Brewer, E., & Halasek, K. (2015). Moving from the one and done to a culture of collaboration: Revising professional development for TAs. *Writing Program Administration*, *39*(1), 32–53.

QDA Miner [Computer software]. (2014). Retrieved from https://provalisresearch.com

Pullin, D. (2008). Assessment, Equity, and Opportunity to Learn. In P. Moss, D. Pullin, J. Gee, E. Haertel, & L. Young (Eds.), *Assessment, Equity, and Opportunity to Learn* (Learning in Doing: Social, Cognitive and Computational Perspectives, pp. 333-352). Cambridge: Cambridge University Press.

Ruggiero, M. A. (2017, April). Remixing responses: How multimodal feedback encourages reflection and awareness. Presentation at Student Success in Writing Conference, Savannah, GA.

Simpson, J. (2017). Responding to our students' writing: What is good for us and for them? *HOW Journal*, *10*(1), 45–52.

Sommers, N. (1982). Responding to student writing. *College Composition and Communication, 33*(2), 148–156.

Sommers, N., & Saltz, L. (2004). The novice as expert: Writing the freshman year. *College Composition and Communication, 56*(1), 124–149. doi:10.2307/4140684

Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication*, *33*(2), 149–183.

Stern, L. A., & Solomon, A. (2006). Effective faculty feedback: The road less traveled. *Assessing Writing, 11*(1), 22–41.

Straub, R. (1996). Teacher's response as conversation: More than casual talk, an exploration. *Rhetoric Review*, *14*(2), 374 - 399.

Straub, R. (1996). The concept of control in teacher response: Defining the varieties of "directive" and "facilitative" commentary. *College Composition and Communication, 47*(2), 223–251.

Straub, R. (2006). *Key works on teacher response.* Portsmouth, NH: Boynton/Cook Publishers.

Straub, R., & Lunsford, R. F. (1995). *Twelve readers reading: Responding to college student writing.* NY: Hampton Press.

White, E. M.& Wright, C. A. (2016). *Assigning, responding, and evaluating writing: A Writing Teacher's Guide*, 5th *ed*. NY: Bedford/St. Martin.

Wordstat [Computer software]. (2014). Retrieved from https://provalisresearch.com

Yoon, D., Chen, N., Randles, B., Cheatle, A., Löckenhoff, C. E., Jackson, S. J., ... & Guimbretière, F. (2016). RichReview++: Deployment of a collaborative multi-modal annotation system for instructor feedback and peer discussion. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 195–205). ACM.

Appendix. Additional Tables

Table A1

MA1 Use of Novice Terms--All Cohorts

MA1	FA12		SP13		FA13		SP14		FA14		SP15		FA15		SP16	
	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2
AWKWARD	0.41%	1.34%	0.55%	1.22%	0.69%	1.42%	1.32%	0.65%	0.64%	0.39%	0.33%	0.57%	0.54%	0.78%	0.54%	0.29%
COMMA	1.87%	1.93%	2.37%	1.31%	2.10%	1.74%	2.79%	1.95%	1.99%	2.26%	1.70%	0.86%	1.24%	1.27%	1.77%	0.00%
CORRECT	1.14%	1.84%	0.97%	1.48%	2.34%	1.74%	2.40%	1.43%	3.24%	1.30%	1.17%	2.30%	1.84%	1.97%	1.30%	0.14%
FLOW	0.44%	0.45%	0.42%	0.17%	0.54%	0.60%	0.54%	0.39%	0.49%	0.48%	0.85%	1.15%	0.51%	0.25%	0.38%	0.43%
GOOD	17.86%	20.91%	16.58%	14.27%	14.92%	18.20%	12.86%	14.69%	20.05%	26.25%	21.59%	23.67%	25.80%	26.61%	17.73%	13.06%
GRAMMAR	0.44%	1.03%	1.31%	1.22%	0.21%	0.33%	0.85%	0.39%	0.56%	0.14%	0.33%	0.43%	0.93%	1.27%	0.23%	0.43%
PUNCTUAT	0.29%	0.67%	0.47%	0.44%	0.24%	0.16%	1.08%	0.65%	0.34%	0.24%	0.07%	0.29%	0.96%	0.49%	0.38%	0.14%
SENTENC	19.06%	17.02%	19.70%	16.28%	15.61%	11.12%	22.08%	14.82%	13.21%	11.66%	15.26%	23.67%	18.11%	13.39%	15.43%	15.35%
SPELL	0.73%	0.76%	0.72%	0.78%	1.26%	1.36%	1.47%	0.65%	0.64%	0.72%	1.17%	1.00%	0.68%	0.21%	0.77%	0.29%
UNCLEAR	0.47%	0.63%	0.89%	1.31%	1.17%	2.67%	0.85%	1.95%	0.71%	1.69%	0.85%	5.60%	1.12%	1.69%	1.53%	3.01%
Semester Avg	4.27%	4.66%	4.40%	3.85%	3.91%	3.93%	4.62%	3.76%	4.19%	4.51%	4.33%	5.95%	5.17%	4.79%	4.01%	3.31%

Table A2

MA1 Use of Expert Terms--All Cohorts

	FA12		SP13		FA13		SP14		FA14		SP15		FA15		SP16	
	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2
ARGUMENT	7.03%	5.73%	7.78%	34.99%	4.68%	6.70%	5.96%	27.96%	10.50%	10.65%	5.41%	40.89%	16.78%	9.55%	8.90%	52.22%
AUDIENC	27.20%	28.53%	1.18%	1.22%	31.33%	35.75%	0.70%	2.08%	33.48%	35.12%	1.24%	2.30%	39.59%	40.22%	1.61%	3.16%
CLEAR	3.75%	5.37%	5.96%	6.35%	3.21%	6.16%	4.34%	7.80%	5.38%	3.81%	6.72%	10.04%	4.68%	6.59%	6.06%	15.21%
COHER	0.12%	0.27%	0.13%	0.35%	0.18%	0.16%	0.23%	0.00%	0.11%	0.05%	0.20%	0.00%	0.12%	0.07%	0.00%	0.14%
DEVELOP	2.84%	3.72%	3.64%	6.88%	2.61%	2.78%	1.86%	2.73%	2.52%	2.26%	4.31%	4.30%	1.77%	1.80%	2.69%	3.01%
EVID	3.95%	4.43%	2.75%	10.97%	5.85%	4.80%	3.64%	9.88%	3.61%	4.43%	2.35%	13.06%	3.80%	3.07%	3.84%	10.62%
FOCUS	2.69%	3.72%	6.22%	1.91%	3.00%	3.76%	4.96%	2.86%	1.92%	2.99%	3.78%	2.58%	3.40%	3.42%	11.28%	8.18%
IDEA	6.15%	4.39%	5.92%	3.39%	4.80%	4.03%	10.61%	5.20%	10.01%	11.37%	10.50%	4.88%	8.58%	6.31%	16.35%	3.73%
ORGAN	2.05%	1.84%	3.42%	4.00%	1.20%	0.71%	3.80%	4.16%	2.07%	2.55%	4.04%	5.60%	1.93%	1.30%	5.45%	1.87%
READER	12.94%	10.70%	6.77%	4.44%	11.22%	10.46%	4.73%	2.73%	7.19%	7.37%	5.74%	4.02%	15.66%	9.69%	16.19%	9.61%
SPECIF	10.80%	8.60%	7.48%	3.57%	7.77%	8.66%	7.98%	5.46%	8.69%	5.78%	7.76%	4.88%	12.68%	8.95%	14.89%	7.46%
STRUCTUR	2.81%	3.13%	2.07%	3.48%	2.73%	2.23%	2.79%	1.56%	2.97%	2.55%	1.70%	3.16%	2.33%	1.16%	1.30%	8.46%
SUPPORT	5.36%	4.12%	3.81%	11.84%	6.93%	6.32%	8.13%	12.22%	3.69%	4.38%	4.44%	8.18%	4.68%	3.74%	6.52%	23.39%
THESI	14.61%	11.15%	11.75%	12.79%	7.17%	5.50%	20.91%	16.38%	8.05%	5.39%	10.18%	22.53%	11.05%	5.92%	17.11%	24.96%
Semester Avg	7.31%	6.84%	4.92%	7.58%	6.62%	7.00%	5.76%	7.22%	7.16%	7.05%	4.88%	9.03%	9.08%	7.27%	8.01%	12.29%

Table A3

Novice Terms--Cohort Comparison

MA1	FA12		SP13		FA13		SP14		FA14		SP15		FA15		SP16		Average
	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2	
AWKWARD	0.41%	1.34%	0.55%	1.22%	0.69%	1.42%	1.32%	0.65%	0.64%	0.39%	0.33%	0.57%	0.54%	0.78%	0.54%	0.29%	0.73%
COMMA	1.87%	1.93%	2.37%	1.31%	2.10%	1.74%	2.79%	1.95%	1.99%	2.26%	1.70%	0.86%	1.24%	1.27%	1.77%	0.00%	1.70%
CORRECT	1.14%	1.84%	0.97%	1.48%	2.34%	1.74%	2.40%	1.43%	3.24%	1.30%	1.17%	2.30%	1.84%	1.97%	1.30%	0.14%	1.66%
FLOW	0.44%	0.45%	0.42%	0.17%	0.54%	0.60%	0.54%	0.39%	0.49%	0.48%	0.85%	1.15%	0.51%	0.25%	0.38%	0.43%	0.51%
GOOD	17.86%	20.91%	16.58%	14.27%	14.92%	18.20%	12.86%	14.69%	20.05%	26.25%	21.59%	23.67%	25.80%	26.61%	17.73%	13.06%	19.07%
GRAMMAR	0.44%	1.03%	1.31%	1.22%	0.21%	0.33%	0.85%	0.39%	0.56%	0.14%	0.33%	0.43%	0.93%	1.27%	0.23%	0.43%	0.63%
PUNCTUAT	0.29%	0.67%	0.47%	0.44%	0.24%	0.16%	1.08%	0.65%	0.34%	0.24%	0.07%	0.29%	0.96%	0.49%	0.38%	0.14%	0.43%
SENTENC	19.06%	17.02%	19.70%	16.28%	15.61%	11.12%	22.08%	14.82%	13.21%	11.66%	15.26%	23.67%	18.11%	13.39%	15.43%	15.35%	16.36%
SPELL	0.73%	0.76%	0.72%	0.78%	1.26%	1.36%	1.47%	0.65%	0.64%	0.72%	1.17%	1.00%	0.68%	0.21%	0.77%	0.29%	0.83%
UNCLEAR	0.47%	0.63%	0.89%	1.31%	1.17%	2.67%	0.85%	1.95%	0.71%	1.69%	0.85%	5.60%	1.12%	1.69%	1.53%	3.01%	1.63%

Table A3 Continued

Novice Terms--Cohort Comparison

MA2	FA13		SP14		FA14		SP15		FA15		SP16		FA16		SP17		
	Draft 1.1	Draft 1.2	Draft 1.1	Draft2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft2.2	
AWKWARD	0.60%	0.90%	0.00%	0.00%	0.42%	0.50%	0.37%	0.33%	1.71%	0.70%	0.28%	0.28%	0.56%	0.21%	0.24%	0.20%	0.46%	decrease
COMMA	1.13%	1.24%	1.44%	0.00%	0.72%	1.09%	1.65%	0.16%	1.71%	0.70%	1.76%	1.41%	0.73%	0.42%	0.57%	0.59%	0.96%	decrease
CORRECT	3.07%	1.47%	0.57%	1.79%	0.66%	1.29%	1.04%	1.47%	2.34%	2.11%	1.53%	2.44%	0.97%	0.63%	2.43%	0.98%	1.55%	decrease
FLOW	0.60%	0.23%	0.00%	0.00%	0.66%	0.40%	0.61%	0.98%	0.54%	1.06%	1.02%	0.28%	0.12%	0.11%	0.57%	0.59%	0.49%	flat
GOOD	15.55%	14.14%	15.52%	25.00%	19.29%	21.69%	15.69%	14.54%	19.64%	28.17%	26.45%	26.36%	19.53%	21.15%	15.36%	15.88%	19.62%	increase
GRAMMAR	0.47%	0.11%	0.57%	0.00%	0.00%	0.20%	0.18%	0.16%	0.83%	1.06%	0.28%	0.19%	6.21%	3.17%	0.00%	0.00%	0.84%	increase
PUNCTUAT	0.27%	0.11%	0.00%	0.00%	0.30%	0.50%	0.06%	0.16%	0.39%	0.44%	0.51%	0.09%	1.67%	0.42%	0.32%	0.20%	0.34%	decrease
SENTENC	17.16%	7.69%	27.01%	17.86%	12.84%	10.05%	18.99%	11.44%	14.56%	13.91%	22.59%	22.05%	17.63%	18.40%	22.31%	10.98%	16.59%	flat
SPELL	0.53%	0.23%	0.57%	0.00%	0.42%	0.50%	0.79%	0.16%	0.39%	0.62%	1.25%	0.84%	1.79%	0.48%	0.49%	0.00%	0.57%	decrease
UNCLEAR	0.27%	0.23%	0.86%	0.00%	1.08%	1.09%	1.10%	2.12%	0.49%	0.97%	0.51%	2.53%	0.64%	0.58%	0.89%	1.57%	0.93%	decrease

Note. Highlighting indicates use at more than 0.10% above mean.

Table A4

Expert Terms--Cohort Comparison

MA1	FA12		SP13		FA13		SP14		FA14		SP15		FA15		SP16		Average
	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2	
ARGUMENT *	7.03%	5.73%	7.78%	34.99%	4.68%	6.70%	5.96%	27.96%	10.50%	10.65%	5.41%	40.89%	16.78%	9.55%	8.90%	52.22%	15.98%
AUDIENC **	27.20%	28.53%	1.18%	1.22%	31.33%	35.75%	0.70%	2.08%	33.48%	35.12%	1.24%	2.30%	39.59%	40.22%	1.61%	3.16%	17.79%
CLEAR	3.75%	5.37%	5.96%	6.35%	3.21%	6.16%	4.34%	7.80%	5.38%	3.81%	6.72%	10.04%	4.68%	6.59%	6.06%	15.21%	6.34%
COHER	0.12%	0.27%	0.13%	0.35%	0.18%	0.16%	0.23%	0.00%	0.11%	0.05%	0.20%	0.00%	0.12%	0.07%	0.00%	0.14%	0.13%
DEVELOP	2.84%	3.72%	3.64%	6.88%	2.61%	2.78%	1.86%	2.73%	2.52%	2.26%	4.31%	4.30%	1.77%	1.80%	2.69%	3.01%	3.11%
EVID	3.95%	4.43%	2.75%	10.97%	5.85%	4.80%	3.64%	9.88%	3.61%	4.43%	2.35%	13.06%	3.80%	3.07%	3.84%	10.62%	5.69%
FOCUS	2.69%	3.72%	6.22%	1.91%	3.00%	3.76%	4.96%	2.86%	1.92%	2.99%	3.78%	2.58%	3.40%	3.42%	11.28%	8.18%	4.17%
IDEA	6.15%	4.39%	5.92%	3.39%	4.80%	4.03%	10.61%	5.20%	10.01%	11.37%	10.50%	4.88%	8.58%	6.31%	16.35%	3.73%	7.26%
ORGAN	2.05%	1.84%	3.42%	4.00%	1.20%	0.71%	3.80%	4.16%	2.07%	2.55%	4.04%	5.60%	1.93%	1.30%	5.45%	1.87%	2.87%
READER	12.94%	10.70%	6.77%	4.44%	11.22%	10.46%	4.73%	2.73%	7.19%	7.37%	5.74%	4.02%	15.66%	9.69%	16.19%	9.61%	8.72%
SPECIF	10.80%	8.60%	7.48%	3.57%	7.77%	8.66%	7.98%	5.46%	8.69%	5.78%	7.76%	4.88%	12.68%	8.95%	14.89%	7.46%	8.21%
STRUCTUR	2.81%	3.13%	2.07%	3.48%	2.73%	2.23%	2.79%	1.56%	2.97%	2.55%	1.70%	3.16%	2.33%	1.16%	1.30%	8.46%	2.78%
SUPPORT	5.36%	4.12%	3.81%	11.84%	6.93%	6.32%	8.13%	12.22%	3.69%	4.38%	4.44%	8.18%	4.68%	3.74%	6.52%	23.39%	7.36%
THESI	14.61%	11.15%	11.75%	12.79%	7.17%	5.50%	20.91%	16.38%	8.05%	5.39%	10.18%	22.53%	11.05%	5.92%	17.11%	24.96%	12.84%

Notes

* 9.18% on assignments not specifically termed "argument"

** 2.10% on 1302 assignments; 1301 were drafts of rhetorical analysis

Table A4 Continued

Expert Terms--Cohort Comparison

MA2	FA13		SP14		FA14		SP15		FA15		SP16		FA16		SP17		Average	
	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2	Draft 1.1	Draft 1.2	Draft 1.1	Draft 2.2		
ARGUMENT *	4.74%	8.26%	6.90%	50.00%	7.84%	6.87%	4.03%	39.54%	8.30%	7.13%	7.26%	30.02%	10.16%	4.49%	8.08%	42.35%	15.37%	decrease
AUDIENC **	30.04%	36.09%	0.00%	0.00%	38.58%	39.40%	0.43%	2.94%	36.83%	36.53%	1.65%	7.22%	48.64%	42.73%	1.21%	7.45%	20.61%	increase
CLEAR	3.87%	5.88%	5.75%	12.50%	3.68%	5.07%	3.48%	5.23%	3.86%	3.26%	9.88%	3.38%	8.23%	9.47%	5.82%	14.71%	6.50%	increase
COHER	0.20%	0.23%	0.00%	0.00%	0.30%	0.20%	0.18%	0.16%	0.05%	0.09%	0.28%	0.09%	0.85%	0.00%	0.00%	0.00%	0.16%	flat
DEVELOP	4.07%	5.09%	4.60%	1.79%	3.74%	2.99%	2.38%	4.74%	2.05%	1.58%	2.21%	9.85%	1.11%	2.59%	1.54%	1.96%	3.27%	increase
EVID	1.94%	4.30%	1.72%	5.36%	5.12%	4.68%	3.60%	9.15%	4.49%	3.61%	2.61%	5.25%	5.21%	3.38%	1.05%	5.49%	4.19%	decrease
FOCUS	3.14%	2.49%	11.21%	3.57%	4.58%	4.18%	9.04%	4.90%	2.20%	1.94%	3.01%	2.16%	3.60%	2.86%	9.94%	2.16%	4.44%	flat
IDEA	4.81%	4.86%	5.46%	5.36%	8.98%	8.66%	6.35%	6.37%	6.01%	7.66%	3.52%	1.78%	4.33%	3.17%	11.48%	2.94%	5.73%	decrease
ORGAN	1.54%	1.47%	4.89%	3.57%	1.99%	1.99%	2.75%	4.08%	1.66%	1.85%	2.89%	0.84%	3.10%	1.69%	2.99%	2.94%	2.52%	decrease
READER	10.41%	9.50%	5.17%	1.79%	5.73%	4.58%	3.05%	4.25%	7.96%	9.07%	5.68%	1.22%	10.48%	10.79%	14.96%	3.73%	6.77%	decrease
SPECIF	7.68%	8.60%	6.90%	12.50%	11.39%	12.44%	8.67%	5.39%	11.43%	10.56%	4.54%	4.78%	15.26%	14.49%	5.25%	3.33%	8.95%	increase
STRUCTUR	1.13%	1.02%	3.45%	3.57%	2.41%	2.39%	2.44%	1.31%	0.78%	1.23%	2.21%	1.22%	1.84%	3.54%	1.78%	0.78%	1.94%	decrease
SUPPORT	2.27%	3.39%	2.59%	8.93%	2.11%	1.59%	2.26%	15.03%	2.88%	3.52%	3.01%	7.22%	4.01%	3.01%	3.96%	11.57%	4.83%	decrease
THESI	7.34%	4.64%	9.48%	14.29%	7.17%	4.48%	12.70%	19.77%	8.06%	3.96%	7.83%	8.54%	15.87%	10.68%	19.00%	17.25%	10.69%	decrease

Notes

Highlighting indicates use at more than 0.10% above mean.

* 7.01% on assignments not specifically termed "argument"

** 2.61% on 1302 assignments