Everything is illuminated: What big data can tell us about teacher commentary

Zachary Dixon a,*, Joe Moxley b

a CPR 304a, University of South Florida, 4202 E Fowler Avenue, Tampa, FL 33620-5550, United States
b CPR 306, University of South Florida, 4202 E Fowler Avenue, Tampa, FL 33620-5550, United States

A R T I C L E   I N F O

Article history:
Received 5 November 2012
Received in revised form 20 August 2013
Accepted 23 August 2013
Available online 19 September 2013

Keywords:
Writing assessment
Eportfolio
Analytics
Big data
Digital rhetoric

A B S T R A C T

What happens to writing instructors’ feedback when they use a common rubric and an online tool to respond to student papers in a first-year composition course at a large state university in the United States? To investigate this question, we analyze the 118,611 comments instructors made when responding to 17,433 student essays. Using concordance software to quantify teachers’ use of rubric terms, we found instructors were primarily concerned with global, substantive, higher-order concerns—such as responding to students’ rhetorical situations, use of reason, and organization—rather than lower-order concerns about grammar or formatting. Given past research has determined teachers overemphasize lower-order concerns such as grammar, mechanics, and punctuation (Connors & Lunsford, 1988; Lunsford & Lunsford, 2008; Moxley and Joseph, 1989, 1992; Schwartz, 1984; Sommers, 1982; Stern & Solomon, 2006), these results may suggest the possibility of a generational shift when it comes to response to student writing. Aggregating teacher commentary, student work, and peer review responses via digital tools and employing concordance software to identify big-data patterns illuminates a new assessment practice for Writing Program Administrators—the practice of Deep Assessment.

© 2013 Elsevier Ltd. All rights reserved.

* Corresponding author. Tel.: +1 407 325 2182.
E-mail address: parke@mail.usf.edu (Z. Dixon).

1075-2935/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.asw.2013.08.002
1. Introduction

Responding to student writing is one of the most time consuming, complicated, and critical responsibilities of writing instructors. Even so, as researchers have remarked over the last century, there’s a gap at times between how we might hope our comments inspire, guide, and educate our students and how we actually respond to writing or how students perceive our comments (Moxley, 1992; Smith, 1997; Sommers, 1982). Studies of teachers’ responses to student writing have repeatedly found that teachers tend to overemphasize local concerns—issues related to “correctness” such as grammar, punctuation, mechanics, and word choice—over more global, rhetorical concerns—issues more related to deep learning matters, such as a student’s ability to summarize multiple sources and put them in conversation with one another from an historical, causal, or analytical perspective (Moxley, 1989, 1992; Connors & Lunsford, 1988; Lunsford & Lunsford, 2008; Schwartz, 1984; Sommers, 1982; Stern & Solomon, 2006).

This research was inspired by the efforts of Nancy Sommers (1982), Robert Connors and Andrea Lunsford (1993), Summer Smith (1997), and Andrea Lunsford and Karen Lunsford (2008) who have endeavored to map “those genres and tropes of response we tend to privilege” (Connors & Lunsford, 1993, p. 219) when we respond to student writing. However, our focus and data collection methods are somewhat narrower than previous researchers. First, unlike the Connors and Lunsford or Lunsford and Lunsford studies, we do not base our research on a large national sample of 3000 or so student papers sent in by well-meaning teachers from around the nation. Instead, our study is focused on the standards of practice of graduate students and adjunct instructors who work at one large writing program at the University, a southern state university in the United States, a program that requires all instructors to employ a shared, standardized curriculum that was developed through the cooperation of the department’s administration, faculty, and graduate teaching assistants. Second, our study is not designed to analyze a sample of papers to deduce the nature of common errors in student work and the instructors’ written response associated with those errors, and then to speculate about the apparent patterns associated with written feedback—a top 20 list of common errors for the nation. Instead our analysis of the 118,611 written comments that our instructors made when grading and responding to 17,433 student essays (which constituted students’ intermediate and final drafts on Projects 1, 2, and 3 in ENC 1101 and Projects 1, 2, and 3 in ENC 1102 during the fall and spring 2011/2012 academic year) was designed to gain insights into the practices of instructors’ written feedback within the context of our own university writing program—a portrait of what written assessment looks like in our own department.

We were especially interested in determining how instructors use our rubric terms to inform their commentary because our program’s rubric was developed through a lengthy and involved crowd-sourcing, effort—what we have described elsewhere as a datagogoical process, a process of making informed, repeated curriculum changes in response to real-time data and dialog among teachers, teachers, and administrators (Moxley, 2008, 2012; Moxley et al., 2012). Because we had gone to great lengths to define these rubric criteria through textbooks, podcasts, videos and wikis (Moxley, 2012; Moxley et al., 2012), we were curious to see if instructors and students explicitly used these terms when discussing texts. In other words, we were interested in exploring how often our instructors explicitly employ language from our programmatic rubric within their own comments because we thought this analysis would give us some insight into our students’ writing and perhaps identify needed additional curricular materials or rubric changes. As writing teachers we prize clarity, and while we recognize that discussions of students’ texts can encompass a wide range of topics—from affirming the student’s efforts to expressing excitement or dismay about the student’s topic, and beyond—we believe that measuring instructors’ explicit use of rubric terms would best give us insights regarding how these terms played out in our community of practice. As such, we hoped to better understand the degree of which instructors explicitly utilize the language from the programmatic rubric in their written responses to student work and to identify significant relationships or patterns between instructors’ written comments and the specific rubric criteria.

Unlike previous studies that relied on the massive effort of large teams of people manually gathering, sorting, reading, and reviewing their data, our study relies almost exclusively on digitally automated processes of collection and analysis. While previous studies of instructor response have
helped to reveal critical flaws involved with instructor feedback, their findings and methodologies have been limited by the time and energy required to assemble, sort, categorize, and analyze a complex and resistant set of data. For example, in Lunsford and Connors’ (1988) first study of student error, they studied approximately 3000 student papers from across the country, sent in from more than 300 instructors. Five years later, Connors and Lunsford’s (1993) follow up was similarly only able to analyze 3000 of the 21,000 essays collected. Summer Smith’s (1997) more recent landmark essay looked at 208 endnote comments collected from ten teaching assistants. Even Lunsford and Lunsford’s (2008) most recent study of error patterns was conducted using only 877 random student papers. Collecting and analyzing a large enough sample of student work and instructor feedback represents a herculean task in terms of time and energy as Connors and Lunsford (1993) note: “we faced the question of what instrument we would use to try to understand what we might find in the 3000-paper sample” (p. 206). New digital research tools, however, enable us to assemble, sort, categorize, and analyze such a complex and resistant set of data quickly and efficiently. Digitized assessment corpuses and automated lexical analysis software now make it possible to do in a few keystrokes what once took years to assemble: a thorough analysis of what instructors are actually writing to their students on a programmatic level.

2. Methodology

At the University, instructors use a tool called My Reviewers to view, evaluate, and respond to intermediate and final drafts of their student’s work. For 1101, the first composition course, students complete three major projects: a bibliographic essay, a thesis-driven essay, and a remediation analysis (see http://fyc.usf.edu for particulars). While the first two 1101 major projects are fairly straightforward essays, the third project is a remediation essay. Students draft an 800–1000 word, thesis-driven reflection about their work remediating their own thesis-driven essay from the second major project into a multimodal digital format. During the time of this study, 1102 students completed three major projects: a rhetorical ad analysis, an essay which utilizes the Rogerian argument format, and a persuasive essay that aims at encouraging direct action into an issue of social significance. In addition to the six reviews by their instructors for the intermediate and final drafts of their work, students also used My Reviewers to conduct approximately nine peer reviews—that is, at least two reviewers per project. All written commentary and numerical grading made by instructors on these six projects were conducted digitally and archived by My Reviewers. By accessing the My Reviewers corpus, we exported 118,611 comments from 17,433 essays stored for the Fall 2011 and Spring 2012, 1101 and 1102 level composition courses.

We used concordance software to analyze all of the instructors’ marginal and endnote comments. While concordance software and student writing corp i have been used in classroom settings in the form of personal corpi to give students a more detailed, reflexive look into the structures of their writing habits, including lexis, grammar, and collocation (Charles, 2012; Flowerdew, 2009; Gilmore, 2009), particularly for L2 writing students (Gaskell & Cobb, 2004; Wu, Witten, & Franken, 2010; Yoon & Hirvela, 2004), they have not yet been applied to a large scale departmental corpus of student and instructor writing.

In order to account for our instructors’ varied rhetorical purposes, the total corpus of instructor comments was then broken down into distinct corpi: a marginal comment corpus; an endnote comment corpus; an intermediate draft comment corpus; a final draft comment corpus; an 1101 comment corpus; 1101 endnote and marginal comment corpi; 1101 intermediate and final draft comment corpi; an 1102 comment corpus; 1102 endnote and marginal comment corpi; and 1102 intermediate and final draft corpi. By dividing the corpus as such we believed that we could more accurately gauge each genre of instructor comments.

To develop our search terms we extricated the key terms from our community rubric, which constitutes the required grading rubric for all Composition 1101 and 1102 instructors (Table 1). To maximize the search potential of each term within the concordance software, and to account for potential vernacular differences between the 61 Fall 2011, and the 47 Spring 2012 instructors represented in the corpus, each search term, where appropriate, was broken down into its core root and given the Boolean search term “*” to account for possible mutations. For example, we broke down the search term
**Table 1**
Community rubric.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Basics</th>
<th>Emerging 0</th>
<th>Developing 2</th>
<th>Mastering 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus (25%)</td>
<td>Basics</td>
<td>Does not meet assignment requirementsa</td>
<td>Partially meets assignment requirements</td>
<td>Meets assignment requirements</td>
</tr>
<tr>
<td></td>
<td>Critical thinking</td>
<td>Absent or weak, thesis; ideas are</td>
<td>Predictable or unoriginal thesis; ideas are partially developed and related to thesis; inconsistent analysis of subject relevant to thesis</td>
<td>Insightful/intriguing thesis; ideas are convincing and compelling; cogent analysis of subject relevant to thesis</td>
</tr>
<tr>
<td>Evidence (25%)</td>
<td>Critical thinking</td>
<td>Sources and supporting details lack</td>
<td>Fair selection of credible sources and supporting details; unclear relationship between thesis and primary and secondary sources/evidence; ineffective synthesis of sources/evidence relevant to thesis; occasionally effective synthesis of visuals/personal experience/ anecdotes relevant to thesis; inconsistently distinguishes between writer's ideas and source's ideas</td>
<td>Credible and useful sources and supporting details; cogent synthesis of primary and secondary sources/evidence relevant to thesis; clever synthesis of visuals/personal experience/ anecdotes relevant to thesis; distinguishes between writer's ideas and source's ideas</td>
</tr>
<tr>
<td>Organization (25%)</td>
<td>Basics</td>
<td>Confusing opening; absent, inconsistent, or non-relevant topic sentences; few transitions and absent or unsatisfying conclusion</td>
<td>Uninteresting or somewhat trite introduction, inconsistent use of topic sentences, segues, transitions, and mediocre conclusion</td>
<td>Engaging introduction, relevant topic sentences, good segues, appropriate transitions, and compelling conclusion</td>
</tr>
<tr>
<td></td>
<td>Critical thinking</td>
<td>Illogical progression of supporting points; lacks cohesiveness</td>
<td>Supporting points follow a somewhat logical progression; occasional wandering of ideas; some interruption of cohesiveness</td>
<td>Logical progression of supporting points; very cohesive</td>
</tr>
<tr>
<td>Style (20%)</td>
<td>Basics</td>
<td>Frequent grammar/punctuation errors: inconsistent point of view</td>
<td>Some grammar/punctuation errors occur in some places; somewhat consistent point of view</td>
<td>Correct grammar and punctuation; consistent point of view</td>
</tr>
<tr>
<td></td>
<td>Critical thinking</td>
<td>Significant problems with syntax, diction, word choice, and vocabulary</td>
<td>Occasional problems with syntax, diction, word choice, and vocabulary</td>
<td>Rhetorically sound syntax, diction, word choice, and vocabulary; effective use of figurative language</td>
</tr>
<tr>
<td>Format (5%)</td>
<td>Basics</td>
<td>Little compliance with accepted</td>
<td>Inconsistent compliance with accepted documentation (i.e., MLA, APA) for paper formatting, in-text citations, annotated bibliographies, and works cited; some attention to document design</td>
<td>Consistent compliance with accepted documentation (i.e., MLA, APA) for paper formatting, in-text citations, annotated bibliographies, and works cited; strong attention to document design</td>
</tr>
<tr>
<td></td>
<td></td>
<td>documentation style (i.e., MLA, APA) for paper formatting, in-text citations, annotated bibliographies, and works cited</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This is a “living document”; in other words, we revise our scoring criteria in response to teacher and student feedback.

a A “0” for “Does not meet assignment requirements” results in an overall 0 (F) for the project.
“formatting” to “format”, which became “format” so that all possible permutations of “format,” such as “formatting,” “formatted,” and “formats,” would be accounted. We believed that although there was overlap in instructors between the two semesters, the possible range of language used by those instructors required the use of the root/Boolean system to ensure an accurate image of rubric language use. In summary, from the FYC rubric, we choose the following terms to represent each of the five critical areas of evaluation:

**Focus**: assignment require*, thesis, analy*, idea*
**Evidence**: source*, evidence, supporting detail*
**Organization**: introduc*, topic sentence*, transition*, conclusion*, logic*, cohesive
**Style**: grammar, punctuation, point of view, word choice, vocab*, syntax
**Format**: MLA, format*, cit*, work* cited

Using the concordance software, AntConc, our search terms were applied to each of our corpi. The primary functions of AntConc that we employed were the “Concordance Tool” and the “Word List”. The AntConc “Concordance Tool” allowed us to apply each search term to each corpus, which returned a respective frequency number. The “Word List” AntConc feature generated a complete list of unique words appearing in each corpus and assigned them a rank based on their frequency.

Given our reliance on digital systems and our specific interest in the use of rubric language in our instructors’ written comments, our methods exhibit a few limitations. First, instructors’ comments are taken out of their natural context. Removed from the student work for whom they were intended and for whom they may not have held value, there is no way to determine the potential their accuracy or helpfulness. That said, while we cannot substantiate the validity of these comments in terms of accuracy or helpfulness we do think they are worth studying for several reasons: (1) they represent literally thousands of hours of reading and responding to student work; (2) they represent the instructors assessments and the instructors are the ones, after all, who assign grades. Additionally, we reaffirm our belief in the wisdom of the crowds: that the more individuals that can contribute to a networked information economy—such as our FYC program—the more enriched the knowledge that is generated as a result of that interaction is (Benkler, 2006; Kelly, 2003). In following with the work of Benkler and Kelly, and in keeping with the general principle of the law of large numbers, we believe that because our instructor’s comments are taken in mass aggregate they represent a valuable articulation of what is privileged in our instructors’ responses to student writing within our program.

A secondary and related limitation of our study design is that our focus on the explicit rubric terminology cannot always cover the details associated with a particular comment. For example:

**Rubric Keywords:**
thesis, introduce*

**Analysis of Instructor Comment:**
Jane, you tell the story well with strong details but your organization is lacking. You don’t have a thesis in your introduction and your body paragraphs are unbalanced; some are well developed; others are underdeveloped.

**Registered rubric terms:**
thesis, introduction

**Unanalyzed Instructor Comment:**
Jane, you tell the story well with strong details but your organization is lacking. You don’t have a ______in your ______ and your body paragraphs are unbalanced; some are well developed; others are underdeveloped.

A third and final design limitation is that the key word ranking is sometimes obscured by superfluous number and letter combinations that are carried over between format changes. For example, when transferring corpus data from .exl spreadsheets into .txt files, where they can be analyzed by AntConc, the alphanumeric sequence “&nbsp;“ is substituted for any non-breaking space. Given the great frequency of non-breaking spaces in a corpus of typed comments, “&nbsp;“ ends up with a key word
3. Results

3.1. Total corpus trends

As illustrated by Fig. 1, of the 118,611 total instructor comments aggregated into the total corpus, comments that directly utilized some variation of the programmatic FYC rubric occurred 76,115 times. Of all the comments, both marginal and endnote, that used language from the community rubric, comments that utilized language from the “Focus” criteria represented the most common form of rubric-oriented comments at 19,971 comments, 26% of all rubric-oriented comments. Within the subset of “Focus” oriented comments, the concepts of “thesis” (7971), “idea” (5750), and “analysis” (3970) were commented on the most. Rubric comments involving “Format” elements were the second most common with 19,297 instances, a quarter of all rubric-oriented comments. Rubric-oriented comments concerning “Organization” and “Evidence” made up the bulk of the remaining rubric comments with 16,017 (21%) and 15,973 (20%) respective occurrences. Only 4857 comments, just over 6% of all teacher comments collected, represented “Style” elements of the programmatic rubric. As a litmus test for the frequency of mechanically orientated comments among the entire 118,611, we found that “comma” represented the 111th most frequent word, far behind “paragraph” at 55th and “thesis” at 68th.

Because each endnote can contain multiple key-words, and because AntConc concordance software only registered each occurrence of rubric key-words, we identified 43,684 rubric-oriented comments within our corpus of 17,480 endnotes (Fig. 2). The 43,684 rubric-oriented endnote comments constitute just over 57% of the total corpus of 76,115 rubric-oriented comments. Again, “Focus” comments represented the most frequent rubric-oriented endnote comment at 12,681 (29%) instances. Those comments which spoke to “Evidence” concerns represented the second most frequent rubric oriented endnote comments, with 10,601 (19%) occurrences. Within rubric-oriented endnotes “Format” and “Organization” oriented comments appeared with similar frequencies, 9547 (21%) and 8340 (19%) times respectively. There were only 2516 rubric language endnote comments geared toward “Style” elements, less than 6% of the total rubric language endnote comments.

Marginal comments, not surprisingly, accounted for the most common form of teacher feedback with 101,131 comments collected. Perhaps more surprising is that only 32,431 marginal comments, or about 32% of all marginal comments and just over 42% of all rubric-oriented comments, exhibited any rubric key-words (Fig. 2). The most common rubric-oriented marginal comment is the “Format” comment with 9750 instances, representing 30% of all rubric oriented marginal comments. With 7677 instances, marginal comments that addressed “Organization” makeup nearly 24% of the total rubric-orientated marginal comments. Rubric-oriented marginal comments that included “Focus”
key-words appeared 7291 times (22%) while rubric-oriented marginal terms that included “Evidence” key-words appeared 5372 times (16%). Again, by far the least frequent rubric-oriented marginal comments were those that included “Style” keywords such as “grammar,” “punctuation,” “point of view,” “word choice,” “vocabulary,” and “syntax,” appeared only 2341 times—representing barely 7% of all rubric-oriented comments. Among marginal comments the phrase “comma*” ranked only 68th in overall frequency, behind both “paragraph*” at 46th and “thesis” at 58th.

### 3.2. Endnote trends

Out of the total comment corpus, 7894 comments represent the endnotes generated for 1101 courses. In the same way that the larger endnote corpus, because of endnotes’ tendency to embody multiple sentences and therefore multiple topics, generated a disproportionate number of rubric-oriented comments, the 7894 1101 endnotes generated 19,568 rubric-oriented comments (Fig. 3). Of the 19,568 rubric-oriented 1101 endnote comments, “Focus” represented the most popular topic with 5503 representative comments; equaling 28% of all the rubric-oriented 1101 endnote comments. Comments that featured rubric language that addressed “Evidence” were the second most frequent rubric-oriented 1101 endnote comments, representing 26.6% of all rubric-oriented 1101 endnote comments with 5221 instances. Rubric-oriented 1101 endnote comments which utilized language from the “Organization” and “Format” sections of the rubric occurred 3682 (18.8%) and 3912 (19.9%) times respectively. Comments which utilized language from the “Style” section of the rubric were the least frequent rubric-oriented 1101 endnote comments, representing only 6.3% of rubric-oriented 1101 endnote comments with 1250 instances.

---

**Fig. 2.** Rubric endnote, \(N = 43,684\); rubric marginalia, \(N = 32,431\).

**Fig. 3.** 1101 endnote, \(N = 19,568\); 1102 endnote, \(N = 21,153\).
The 9586 1102 endnote comments collected by My Reviewers generated 21,153 rubric-oriented comments (Fig. 3). Unsurprisingly, “Focus” once again represented the most frequently utilized rubric language section, accounting for 30.4% of all rubric-oriented 1101 endnote comments with 7054 representative comments. The 5380 comments which utilized language from the “Evidence” section of the rubric made up 23.2% rubric-oriented 1102 endnote comments. Rubric-oriented 1102 endnote comments which featured “Format” language occurred 4850 times (20.9%), while rubric-oriented 1101 endnote comments which featured “Organization” language occurred 4652 times (20%). Rubric-oriented 1102 endnote comments which were concerned with “Style” once again appeared the least, respectively, representing only 5.2% of all rubric-oriented 1102 endnote comments with 1216 instances.

3.3. Marginal comment trends

From the 44,214 marginal comments collected from 1101 course sections, 15,120 exhibited rubric-oriented language (Fig. 4). Of the 15,120 rubric-oriented 1101 marginal comments, in a departure from most of the other data sets, “Format” represented the most frequent comment topic, totaling 30.4% of all rubric-oriented 1101 marginal comments with 4611 comments. The 3517 “Focus” oriented comments were the second most frequent rubric-oriented 1101 marginal comments, representing 23.2% of all rubric-oriented 1101 marginal comments. Comments which exhibited language from the “Organization” section of the rubric were the third most frequent, account for 20.8% of rubric-oriented 1101 marginal comments with 3146 instances. The 2797 “Evidence” based comments represented 18.4% of rubric-oriented 1101 marginal comments, while rubric-oriented 1101 marginal comments which involved “Style” totaled only 1049 for 6.9%.

The 56,917 marginal comments collected from 1102 courses yielded 16,785 rubric-oriented 1102 marginal comments. Similar to rubric-oriented 1101 marginal comments, the most frequent rubric-oriented 1102 marginal comments, representing 27.5% with 4628 instances, used language from the “Format” section. “Organization” was the second most frequently utilized subject for rubric-oriented 1102 marginal comments, representing 26.9% of rubric-oriented 1102 marginal comments with 4521 instances. Comments which utilized language from the “Focus” and “Evidence” sections of the rubric represented 22.4% and 15.3% of rubric-oriented 1102 marginal comments respectively. The least frequent rubric-oriented 1102 marginal comments were once again those which utilized language from the “Style” section of the rubric, counting it at 1289 comments for only 7.6%.

3.4. Intermediate draft trends

Of the 118,611 total instructor comments collected, 71,940 comments were collected from the intermediate drafts of the 1101 and 1102 projects (Fig. 5). Out of the 71,940 intermediate draft
instructor comments 41,340 comments, more than 57% of all intermediate draft instructor comments, exhibited rubric language. Surprisingly, comments which demonstrated language from the “Format” section of the rubric were the most frequently occurring intermediate draft rubric comments, representing 26% of all intermediate draft rubric-oriented comments with 10,779 instances. Intermediate draft comments from the “Focus” section of the rubric, accounting for 24%, of all the intermediate draft rubric-oriented comments, were the second most common with 9927 instances. The 9141 intermediate draft rubric-oriented comments which exhibited language from the “Organization” section of the rubric narrowly surpassed the 8992 intermediate draft rubric-oriented comments which exhibited language from the “Evidence” section; each accounting for 22.1% and 21.7% of all intermediate draft rubric-oriented comments respectively. Like many of other groupings, language from the “Style” section of the rubric once again represented the least common comment within the corpus of intermediate draft rubric-oriented comments with only 2501 instances accounting for 6%.

Instructors from the 1101 courses generated 34,010 comments on their students’ intermediate drafts, 19,633 of which exhibited rubric language (Fig. 6). Reflecting the larger trend for the total corpus of intermediate draft rubric-oriented comments, 1101 intermediate draft rubric-oriented comments utilized language from the “Format” most frequently with 5691 instances; accounting for 28.9% of all 1101 intermediate draft rubric-oriented comments. The 23.7% of 1101 intermediate draft rubric comments that used language from the “Evidence” section of the rubric made it the second most referenced category with 4671 comments. The 4485 1101 intermediate draft comments that referenced the “Focus” section of the rubric accounted for 22.8% of all 1101 intermediate draft rubric-oriented

![Fig. 5](image5.png) **Fig. 5.** Intermediate draft rubric comments, N=71,940.

![Fig. 6](image6.png) **Fig. 6.** 1101 intermediate, N= 19,633; 1102 intermediate, N= 37,930.
comments, making it the third most frequently cited rubric section. Finally, the “Style” section of the rubric once again was the least frequently utilized rubric section with only 1146 for 5.8%.

Instructors from the 1102 courses generated 21,731 comments which utilized rubric language from the entire set of 37,930 1101 intermediate draft comments collected. The most common 1102 intermediate draft rubric-oriented comments, representing 25.4% of all 1102 intermediate draft rubric-oriented comments, were those which utilized language from the “Organization” section of the rubric with 5531 instances. “Focus” was the second most cited section of the rubric for 1102 intermediate draft rubric-oriented comments with 5442 references for 25%. Representing 23.3% of all 1101 intermediate draft rubric-oriented comments, the “Format” section of the rubric was referenced 5082 times. “Style” was once again the least referenced section of the rubric with only 1355 occurrences in 1102 intermediate drafts.

3.5. Final draft trends

The final drafts from both 1101 and 1102 courses generated 46,671 instructor comments, with 32,644 of those instructor comments exhibiting language from the programmatic rubric (Fig. 6). Comments that exhibited language from the “Format” section of the rubric were the most popular final draft rubric-oriented comments, representing 26% of the final draft rubric language corpus with 8520 occurrences. The 7954 comments using language from the “Focus” section of the rubric represented 24.3% of final draft rubric-oriented comments. The 6981 final draft rubric-oriented comments which exhibited language from the “Evidence” section of the rubric narrowly surpassed the 6839 final draft rubric-oriented comments which exhibited language from the “Evidence” section; each accounting for 21.3% and 20.9% of final draft rubric-oriented comments respectively. With only 2350 references, “Style” was the least common final draft rubric-oriented comments.

Instructors from the 1101 courses generated 18,098 comments on the final drafts of their students’ work, 14,790 of which—more than 81% of their respective corpus—directly utilized language from the rubric (Fig. 7). The 3615 comments that utilized language from the “Format” section of the rubric represent the most popular 1101 final draft rubric-oriented comments, account for 24.4% of all 1101 final draft rubric-oriented comments. Representing 23.5% of 1101 final draft rubric-oriented comments, the “Focus” section of the rubric was referenced 3479 times. “Evidence” was the third most referenced section of the rubric for 1101 final draft rubric-oriented comments, accounting for 22.6% of 1101 final draft rubric-oriented comments with 3348 instances. The 3193 “Organization” based comments represented 21.85% of 1101 final draft rubric-oriented comments, while 1101 final draft rubric-oriented comments which involved “Style” totaled only 1155 for 7.7%.

The 28,573 comments collected from 1102 courses’ final drafts yielded 17,782 rubric-oriented comments (Fig. 8). The most frequent 1102 final draft rubric-oriented comments, representing 27.5% with 4905 instances, used language from the “Format” section. “Focus” was the second most frequently
utilized rubric section for 1102 final draft rubric-oriented, representing 25.1% of 1102 final draft rubric-oriented with 4475 instances. Comments which utilized language from the “Organization” and “Evidence” sections of the rubric represented 20.6% and 19.8% of 1102 final draft rubric-oriented respectively. The least frequent 1102 final draft rubric-oriented comments were once again those which utilized language from the “Style” section of the rubric, counting it at 1195 comments for only 6.7%.

4. Discussion

Nearly every aspect of daily life is becoming digitized: our purchases, preferences, and searches. In aggregate, our digital traces shape a profile of who we are, what we may desire/buy, or even whom we may become. From Google’s use of previous search terms to develop suggestions for future searches, to law enforcement agencies use of facial recognition and license-plate reading technologies to track the movement of those they investigate, to even the National Security Administration’s use of cellular phone data to track the behavior of global citizens, corporations and governments are beginning to exploit the massive and ever growing oceans of data. Likewise, the act of composing is being translated into aggregable and digestible bytes of information. As composing takes place in digital spaces, researchers are beginning to employ data-mining strategies to apply rigor and depth to composition research (Vieregge et al., 2013; Lang & Baehr, 2012). When students conduct their writing in digital formats their favorite subjects, their stylistic traits, their common errors, their strengths, and their instructor’s responses become available for analysis. The more students write in digital spaces, the more information becomes available.

When the bytes of student writing are aggregated together into digital corpi like My Reviewers and when those corpi are analyzed using tools like AntConc they become more than a sea of information, they become a resource that supports context based research that is quick, valid, and actionable. We view the digital analysis of large writing corpi as a move toward what we call Deep Assessment. Our metaphor of depth stems from the way that such aggregation and analysis helps researchers peer below the surface of classroom activities into a sort of back-channel of student and instructor practices that are normally hidden. Digital corpi, tools, and methods—the cornerstones of Deep Assessment tools—open access to that back-channel, provide writing researchers avenues for conducting research and obtaining proof points that would normally be impractical or even unattainable otherwise (Fishman, 2012; Haswell, 2012). Not only do Deep Assessment tools and practices provide access to these rich research resources, they also allow researchers to navigate those resources quickly and efficiently.

Instead of teams of researchers conducting work over the course of years, digital affordances expedite the collection and analysis process exponentially. The collection of data for Connors and Lunsford’s seminal 1988 work on written responses to writing required relying on the graciousness of
instructors from around the nation (1988) and took 3 months to accomplish (Lunsford & Lunsford, 2008). Data collection for Lunsford and Lunsford's recent study took a year and a half (2008). The 21,000 papers collected Connors and Lunsford's original survey had to be pared down to a manageable 3000 representative papers because of the daunting amount of work that reading, coding, and analyzing such a large sample represented (1988). For their 1993 follow-up study Connors and Lunsford enlisted the aid of 26 readers to help them sort through their corpus which had been collected for their earlier study in search of only "global comments" (1993). The present study's analysis of 17,433 student papers, because it utilized digital affordances, was conducted by two researchers over the course of one academic semester.

The ease with which collection and analysis can be done using digital tools represents an extraordinary benefit; if a process is made easier to accomplish the more likely that process is going to be accomplished (Benkler, 2006). The value-added efficiency and reduced workload that Deep Assessment tools and practices enable means that writing researchers can conduct more expansive and in-depth studies faster, more frequently, and with less expenditure. This means that the results generated from such studies are likely to more closely reflect the trends and contexts of the writing analyzed, which can be temporally sensitive. For example, Smith's (1997) study relied on instructor comments written in response to student writing done more than ten years earlier (1997). While her study was clearly conducted with the highest quality, and while her results are unquestionably valuable to the field, the delay between writing and analysis is problematic; particularly when one considers the pace of change in other aspects of life during that time period. It is not unreasonable to imagine that in the decade that saw the fall of the Berlin Wall, the end of the Cold War, and the entire Gulf War, and the rise of the internet, writing and written assessment practices may have also changed. Deep Assessment tools and practices create the opportunity to reduce lag time between data generation, collection, and analysis dramatically. The corpi in this study were generated and collected in real-time, and analyzed during that academic semester by writing program administrators eager to make evidence-based curriculum changes. Deep data methods more accurately reflect the content and contexts of writing that is happening now. In terms of writing administration and assessment this means that because digital tools make the analysis of assessment policy and practice easier to accomplish, the more likely that analysis is to take place in a timely and efficient manner, and the more likely that the results of such analysis will generate valid and meaningful conclusions that benefit their writing community.

When a writing community’s network is augmented with digital tools, its ability to produce and implement practical pedagogies are greatly improved because the real-time review and response capabilities that digital tools can provide allow for an increased sensitivity to the local contexts, rhetorics, and characteristics of writing that represent the most valuable sites of assessment (Moxley, 2008; Rice, 2011). Because digital corpi collect data generated by students writing for and within their respective writing programs, research done on that data is able to retain the meaningful context that shapes the subjects, purposes, and meanings of that writing. The sensitivity to writing contexts that Deep Assessment tools and practices lend to the research process make them ideal for generating teacher-based research, investigations of writing program lore, and institutional critique (Nickoson, 2012; North, 1987; Porter et al., 2000).

The present study convincingly demonstrates that within our writing community instructors in our program comment more frequently on macro-level rhetorical concerns, utilizing the rubric section language of “Focus,” “Organization,” and “Evidence,” rather than on the sentence-level concerns epitomized through the “Style” section of the rubric. That the instructors chose comments that invoked “paragraph” 11,025 times in comparison to the 4200 times “comma” was chosen is exciting. Even among marginalia, where one might expect sentence-level comments to proliferate, “paragraph” continues to trump “comma” 5907 times to 3684. These results may suggest that the current generation of composition instructors at our instruction has moved beyond the shortcomings in responding to writing noted by the work of scholars (Moxley, 1989, 1992; Connors & Lunsford, 1988; Lunsford & Lunsford, 2008; Schwartz, 1984; Sommers, 1982; Stern & Solomon, 2006).

Some of the perpetual criticisms of instructor feedback were also evident in our study’s analysis. That instructors for both 1101 and 1102 courses disproportionately commented on elements from the “Format” section of the rubric on both drafts of students’ work, itself worth only 5% of a student’s project
grade, signifies an intriguing issue. While the results of this study tell us little about why instructors disproportionately comment with language from the “Format” section about the rubric, or even how and where those comments are being directed, they do indicate a programmatic issue that we are anxious to explore in greater detail. Because our corpus captures the work of students and instructors as they compose in the specific discourses of our program, we know that the disproportionate use of “Format” rubric language is not just a nebulous issue of written assessment but rather an issue that is likely intimately related to the institutional practices of our program and, as such, provides meaningful reflections on the context that generated them.

The results of our research into the products of our program’s context provide us useful insights into how we might augment that context. Since the beginning of our program we have adopted the same attitude as Star Trek’s United Federation of Planets’ so called “Prime Directive”: do not interfere with the natural development of others. We believed that instructors, if left to their own devices, would develop their own best practices that would satisfy institutional and departmental expectations for providing students with quality responses to their writing. But after reviewing the corpus of their work, and the wide discrepancy between those whose commentary closely matches how we hope the communal rubric might be used, and those whose commentary does not, we have come to question whether a more hands on approach is necessary. We intend to explore the values associated with the rubric itself, the curriculum designed around the rubric, and instructor training in how to use the rubric in order to understand why instructors are commenting so heavily on the “Format” section of rubric, and whether action should be taken to change instructors’ written assessment practice. Perhaps the answer is as simple as reducing the percentage value of the “Style” section of the rubric in order to increase the value of “Format,” thus better matching the rubric’s values to the instructors’ use of it.

Following the premise of the law of large numbers, a basic mantra of Deep Assessment is that the more data one can generate and analyze, the more valid the conclusions extrapolated from that data will be. When one’s conclusions are more valid, they are more actionable. Analyzing a few of our instructors’ comments would tell us very little about the nature of written commentary in our entire program; at the very best such a study could only help improve the responses of the individual instructors whose work was selected. However, when we are able to gather, organize, and study the whole of written responses generated within our program we are able to gain insights into the larger patterns of the entire program’s activities, and through those insights formulate responses that may help correct or laud the practices of our instructors. That nearly every corpus analyzed displayed the disproportional referencing of the “Format” section of the rubric, and not just a few instructors or specific course levels or even specific writing projects, makes it compelling evidence of a programmatic issue; and because it is compelling, that evidence is more likely to affect a change in curriculum or training further down the road. By allowing researchers easy access to programmatic specific data at an institutional level, Deep Assessment can provide opportunities for instructor training, for mentoring, and for assessment practices by allowing administration and researchers to apply greater scrutiny to traditional policies and practices assumed to be affective.

A particularly poignant example of how the analysis of assessment data with digital tools can help augment a writing program is seen where, despite dynamic shifts in the practice of composition toward more open and public practices in the Web 2.0 world, the disciplinary practices of Composition and Rhetoric, and of Writing Assessment, remain largely private endeavors (Smith, 1997). The classroom practices of composition instructors, writing response and assessment included, still take place largely behind closed office doors. Rarely do instructors share their comments or their methods with their colleagues, or even review their products and methods themselves (Smith, 1997), meaning that best-practices aren’t learned or taught, thus not perpetuated. Deep Assessment reveals how our instructors are interacting with their students writing behind closed doors, in spaces which are normally outside of the purview of administration. Because we can see our instructors’ written responses more easily and clearly, we feel that we are in a better position to help guide instructors toward best practices. In the near term we plan on using our study results as a learning tool in our new instructors’ orientation, educating them on what we observe as the normal practice in our institution and helping them understand our expectations about what we would hope from their future interactions with students. We also plan to augment our new instructor mentor program, asking mentors to workshop with their mentees on written responses to their students’ work.
Although we recognize the complicated ethical implications of collecting and mining such data, we stress that when Deep Assessment research, and the policies that could potentially develop as a result of that research, are guided by the principles of an open community of learning that it represents a positive enterprise (Moxley, 2008, 2012; Moxley et al., 2012; Vieregge et al., 2013). Our use of Deep Assessment may sound as if we are constructing a Composition Panopticon, where we see and hear all that is said and done within our program, but we believe that a panopticon isn’t pernicious so long as it’s guided by the principles of the community of learning rather than the community of power (Moxley, 2008, 2012; Moxley et al., 2012). 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.

5. Conclusion

As a space in which hypotheses can be generated, assumptions tested, and conclusions generated, the intersection of digital corpi and digital analysis promises to add new tools and practices to our research repertoire. We believe that Deep Assessment represents a crucial step in the development of an ecological model of writing assessment (Wardle & Roozen, 2012) that better conceptualizes the true breadth and depth of composition’s practice by gathering the input of students and instructors “over time and at multiple locations” (Wardle & Roozen, 2012). The breadth and depth of the research opportunities embodied by such a space are so great at the present study clearly scratches at but one finite aspect of the research and knowledge possible.

We also recognize that the kinds of relatively simple quantitative data generated by the current study will not answer all questions. In order to produce knowledge for our discipline the use of digital corpi and digital analysis tools will demand the continued application of the hermeneutical practices and skills already available: only with powerful critique and narrative will the statistics generated by digital tools come to gain significant meaning. Given the immensely complicated and fluid nature of composition purely quantitative methods are unlikely to ever push the discipline into reducing research subjects to simple discrete and comparable units (North, 1987). In fact, as Haswell and Fishman both recognize, quantitative results are likely to only confirm writing as a complex, often contradictory act, instead of simplifying it (2012, 2012). Coupled with qualitative research methods and careful research designs, the kinds of quantitative research generated using digital tools provide a powerful addition to the development and legitimation of our knowledge of writing.

The use of large-scale digital corpi and analysis tools also opens new ethical dimensions which are, as governments and corporations are also encountering, complex and potentially dangerous. Even a corpus as seemingly benign as classroom assignments and their evaluations carries with it a great deal of potential evils for students and instructors alike. It does not take a wild imagination to envision the potential abuses that could result from a record of every written word, every error, and every opinion, as the temporary becomes the archived and searchable. As Deep Assessment practices develop it is our imperative to remain particularly sensitive to the complex issues of identity and privacy, and their protection. What data is appropriate for collection and analysis, who has access to and control of that data, and how that analysis might be conscientiously done are at the forefront of current debates. Given our intimate humanistic roots, writing studies occupies potentially pivotal and meaningful location for helping influence the ethics of mining the oceans of digital data.

Recognizing that not every writing program is in a position to purchase or develop the kinds of digital tools discussed in this study our program is working hard on making our corpi available to our own instructors and graduate students as well as those from other universities from around the country in order to open our own data to the imagination of others with the hope of spurring research opportunities and questions we may have not considered. We envision an open community of researchers that share their corpi, or who contribute into a single massive composition corpus, in order to help map out the common patterns in writing across multiple institutions, communities, and meanings; creating a sort of general litmus test for who, what, when, where, and how in writing that is available for the entire discipline.
Beyond investigating the nature of instructor commentary we are particularly curious about the nature of peer response to student writing. Do students interact with writing the way their instructors do, or the same as each other? The uniformity of digital information—it is all just code after all—creates opportunities to explore subjects like the nature of graded assessment in new ways. We can imagine, for example, studies that utilize years of information to study specific student groups longitudinally, investigating the relationship between different student and instructor groups—of different gender, ethnicity, primary language, skill levels, even GPAs—to their writing and to each other. The ability of digital corpi to easily incorporate new data, the longevity of that data, and the ease at which that data might be accessed and processed could help us investigate the nature of transfer and of learning: Do students incorporate the advice of their instructors or peers? How much do students alter their work from one draft to the next?

The degree to which new research methods and ideologies are imported into Writing Studies (North, 1987) represents one of discipline’s great strengths, and is part of what allows us to remain innovative and brave in our work. As a result of that innovation, the idea of making monolithic claims about the act of writing based on a qualitative analysis of a handful of students strike us as deeply anachronistic. Likewise, in the future, the notion that we could make definitive claims about acts of writing based on a sample of 3000 students work done out of academic context is likely to strike us as equally antiquated. Applying digital tools and research methods the vast sea of digital data that is being generated by our students and instructors can provide us with millions of data-points; each one a tiny light shining on a previously dim—potentially unimagined—aspect of the composing experience. By investing our attention and energy into making the best possible use of these digital spaces and tools we can continue to remain brave, innovative, and relevant.

References


Nickoson, L. (2012). Revisiting teacher research. In L. Nickoson, & M. P. Sheridan (Eds.), Writing Studies Research in Practice: Methods and Methodologies (pp. 185–196).


Zachary Dixon is a doctoral student at the University of South Florida studying the intersection of digital assessment tools and pedagogy, Sophistic rhetorics, and the applications of Actor Network Theory to Rhetorical Theory.

Joe Moxley, [http://joemoxley.org](http://joemoxley.org), is a professor of English and director of first-year composition at the University of South Florida. Moxley has published books and articles on assessment, learning analytics, and commons-based peer production. Currently, he serves as executive editor and publisher of *Writing Commons*. 