Assessing Peer and Instructor Response to Writing:
A Corpus Analysis from an Expert Survey

Feb. 7, 2017

Ian G. Anson*
Chris M. Anson†

Abstract: Over the past 30 years, considerable scholarship has critically examined the nature of instructor response on written assignments in the context of higher education (see Straub, 2006). However, as Haswell (2005) has noted, less is currently known about the nature of peer response, especially as it compares with instructor response. In this study, we critically examine some of the properties of instructor and peer response to student writing. Using the results of an expert survey that provided a lexically-based index of high-quality response, we evaluate a corpus of nearly 50,000 peer responses produced at a four-year public university. Combined with the results of this survey, a large-scale automated content analysis shows first that instructors have adopted some of the field’s lexical estimation of high-quality response, and second that student peer response reflects the early acquisition of this lexical estimation, although at further remove from their instructors. The results suggest promising directions for the parallel improvement of both instructor and peer response.

Keywords: Peer response, peer review, teacher response, corpus analysis

Highlights:

- A major survey generated a lexicon of terms favored in “principled” response to student writing and a lexicon of terms expected to be used by novices.
- The most frequent terms were applied to a corpus analysis of approximately 100,000 teacher responses and student peer reviews.
- Results show that instructor response weakly meets the standards of “principled response.”

1 This is a pre-print version. For the published version see Anson, Ian G., and Anson, Chris M. “Assessing Peer and Instructor Response to Writing: A Corpus Analysis from an Expert Survey.” Assessing Writing 33(1): 12-24.
* Assistant Professor, UMBC Department of Political Science. 1000 Hilltop Cir., 305 PUP, Baltimore MD 21250. iganson@umbc.edu.
† Distinguished University Professor and Director, Campus Writing and Speaking Program, North Carolina State University. Box 8101, North Carolina State University, Raleigh, N.C. 27695. chris_anson@ncsu.edu.
Student peer responses, however, were not substantially different from instructor response.

When responding to written work, do teachers use preferred practices? Do their students learn and model those practices? Scholars from a variety of disciplines have investigated the quality and content of instructor response to writing, often concluding that instructors focus their responses on superficial or “lower-order” concerns such as grammar, spelling, and wording at the expense of more complex rhetorical, structural, and meaning-based considerations (e.g., Connors & Lunsford, 1988). Some recent work has offered more reason for optimism, arguing that instructor response may be undergoing a “generational shift” toward higher-order considerations by virtue of scholarship in writing studies and writing-across-the-curriculum initiatives (Dixon & Moxley, 2013). By higher-order, we refer to feedback that assesses broader, conceptual-level features of writing, such as “the development of ideas, organization, and the overall focus” of a text (Keh, 1990, 296; see also Nystrand, 1984). However, these developments have yet to influence practice across a wide variety of higher education contexts (e.g., Bailey & Garner, 2010).

In this study, we extend what is known about these questions by examining the properties of teacher and peer response on drafts of writing assignments. As a strategy for improving metacognition, revision and editing, awareness of audience, and other crucial skills, peer response represents an increasingly popular method used by writing instructors (e.g., Berg, 1999; DiPardo & Freedman, 1988; Lundstrom & Baker, 2009). We first seek to refine the prevailing understanding of response by establishing a lexicon of response terms from a national survey of experienced writing instructors and scholars. Respondents in this study value response that communicates principles of audience and purpose, argumentation, clarity and cohesion, and
evidence—principles that, when understood to apply to specific genres, contexts, and norms, are often considered to be the bedrock of successful writing.

We next compare these expert conceptual judgments of principled response to the empirical realities of instructor and student commentary. Through methods of automated content analysis that allow for a detailed examination of the distributions of key terms, we study the content of large corpora of both instructor and peer comments from introductory composition courses at a four-year public university with a well-managed first-year writing program. Results of this analysis show, as might be predicted, differences in the adoption of the key terms between instructors and students. However, these differences are less extreme than might be expected; while instructors include some aspects of principled response in their comments, peer response does as well, although to a lesser extent. In the context of a large, nationally representative, well-informed writing program that relies on state-of-the-art technology for content management, the study offers promise for the implementation of common principles of response—as manifested in the use of threshold writing concepts (Adler-Kassner & Wardle, 2015)—across both instructor and student peer cohorts.

**Instructor Response: Theory and Method**

For decades, scholars have lamented the dearth of high-quality response to student writing at the college level (e.g., Bailey & Garner, 2010; Duncan, 2007; Lunsford & Lunsford, 2008; Stern & Solomon, 2006). Written feedback on student writing is often found to be unhelpfully terse, focusing on local issues such as grammar, sentence construction, and word choice, rather than on higher-order aspects of meaning and broad discursive elements of
students’ texts. The marginal-comment format of written feedback leads instructors to “mark up” papers and insert brief snippets of locally-focused commentary instead—a practice introduced into the first freshman writing program at Harvard at the turn of the 19th century (Author, 1989).

In a series of interviews with instructors in the United Kingdom, for example, Bailey and Garner (2010) found that instructors often feel as though their own commentary is substantially lacking. However, because of large class sizes and teaching loads, the same instructors acknowledge that providing high-quality written response is often unrealistic. These perceptions have been echoed across university contexts in the United States and other regions of the world (Ferris et al., 2013; Montgomery & Baker, 2007).

If instructors acknowledge that their response to student writing is often lacking, what constitutes high-quality response? While the previous studies make it clear that response should encompass a variety of concerns, Stern and Solomon (2006) build on these foundations to argue that instructor response should be like a “road map”, highlighting the present status and future trajectory of students’ emerging texts (e.g., Connors & Lunsford, 1988; R. Lunsford, 1997).

However, some scholars express concern that avoiding grammar and sentence-level minutia by commenting holistically on broad issues may not be especially helpful. Straub (1977), for example, argues that high-quality response should ideally target “middle-level” considerations, which address the quality of specific thoughts and claims, the procedure or techniques associated with the development of ideas or concepts, support and evidence for those ideas, content clarification, and paragraph structure and style. In large-scale studies of response, a number of scholars identify these considerations as rarely commented upon by instructors (Dixon & Moxley, 2013; Harris, 1992; Straub, 1997; 2002).
In addition to the level of commentary, writing scholars have also pointed to the tone of comments as an important feature of their quality. Striking a balance between critique and praise helps to avoid overwhelming students and encourages them to exert further effort to improve their writing (Nelson & Schunn, 2009; Stern & Solomon, 2006). Positive affect can motivate students to engage in good-faith revision efforts, whereas negative tone in response to writing can also precipitate threat responses that derail motivation to learn from shortcomings or effectively revise a draft. In evaluative situations, students’ learning is often forestalled in proportion to how strongly they are dealing with the interpersonal dynamics of potential “face threats” (Authors, 2016). To be effective, instructors must strike a delicate balance as they formulate their responses, often causing them to invest large amounts of time in the process (Author, 1999; Duncan, 2007).

Another perspective on response invokes the concept of self-regulated learning. Nicol and Macfarlane-Dick (2006) argue that while response should relate to specific targets, criteria, and other external reference points, successful students will set internal goals and monitor or regulate their behavior in order to improve their drafts. From this perspective, response to writing should consider how best to cultivate these self-regulating processes, first by explaining what constitutes a successful performance at the outset, as part of an assignment (see Authors, 2015). When providing response, instructors should facilitate self-assessment by encouraging dialogue and self-reflection, calling attention to specific features of writing, and bolstering students’ self-esteem, among other best practices. These aspects of response are likely to improve students’ ability to link external and internal goals, motivating them to take action toward improving written drafts. Instructors are encouraged to provide detailed, friendly, constructive response that limits itself to just a few major points of emphasis (Lunsford, 1988).
Beyond these general guidelines and practices from individual scholars, specific features of high-quality response are less known (see Author, 2012). Presumably, certain important underlying concepts are embedded in, and invoked through, responders’ choice of language and terminology when commenting on what they see in emerging drafts. Determining what constitutes this language in the collective view of many professionals in the field, and then examining how or whether it appears in both instructor and peer response to writing in progress, can offer important insights for teacher development and the implementation of peer-response practices in the classroom.

**Peer Response and the Question of Expertise**

In recent decades, facilitating interactions between and among students in academic settings has come to be considered a significant “high-impact” practice (Salomon & Perkins, 2009). In the context of writing across the disciplines, the use of peer response for students to provide feedback on each other’s drafts in progress has received considerable attention and advocacy in the instructional literature (e.g., Cho, Schunn & Charney, 2006; Leijen & Leontjeva, 2012; Paulus, 1999). Especially popular in L2 instruction, peer response has been shown to help students understand their own process of writing development by analyzing the writing of peers at similar stages in the process. It has been integrated into a number of theoretical approaches to writing instruction: in early process writing theory (Elbow, 1973), peer response was said to improve a novice writer’s understanding of audience and the perspective of the reader, which has been shown to improve the ability to detect and diagnose writing problems (Flower & Hayes, 1981). In models based on collaborative learning theory (CLT), as students provide response to
their peers, they are expected to benefit from the guidance of other peers who have commented on their work in a reciprocal fashion (Bruffee, 1984; Hansen & Liu, 2005).

However, the content of peer response can critically influence its effectiveness. Becker (2006), for example, claims that revision is often practiced by novices as a “red pen”-style task of surface-level correction, meaning that these students provide superficial commentary on drafts when instructors do not properly introduce, explain, and model the method. Students responding in this manner may not grasp the full importance and utility of more robust commentary, even after multiple rounds of revision (Flower, et al., 1986). In contrast, those instructors who have incorporated goal-oriented training into their implementation of peer response practices can help students to focus on specific, principled aspects of the writing task (Leijen, 2014). Yet studies have found conflicting results for the effectiveness of peer response across the implementation of specific strategies and practices (e.g., Goldin, Ashley, & Schunn, 2012; Hayes, 2012). Notably, Comer et al. (2014) examined peer-to-peer feedback in the context of large-scale MOOCs, finding that higher- and lower-order concerns in peer feedback varied across disciplines.

Despite numerous advances in the study of peer response, these and other uncertainties in the research point to a need for a fuller understanding of whether the content of peer response reflects principled approaches. In particular, research on the relationship between teacher response and peer response in the same courses is lacking. 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?
Study 1: An Expert-Driven Approach to Assessing Response

To answer these questions, we report on a three-stage study. First, we refine the definition of “quality feedback” through a lexically-based instrument created from a survey of experienced instructors, scholars, and program administrators. Second, we introduce and analyze the content of a large repository of instructors’ response to student writing. Finally, we compare this analysis to a study of peer response generated about drafts of the same student papers commented on by their instructors. Unlike studies that relate specific kinds of peer response to specific revisions (or lack thereof) in writers’ drafts (e.g., Leijen, in press; Nystrand, 2002), the present study focuses on the conceptual language teachers expect will be used in expert vs. novice response to student writing, and the extent to which that conceptual language appears in actual teacher and peer response to students’ texts.

In 2016, we posted an invitation to participate in a non-probability, voluntary survey of writing experts on the listserv of the Council of Writing Program Administrators, WPA-L. This listserv is populated by approximately 3,500 members, almost all of them with expertise as teachers of college-level writing courses and/or administrators of university expository writing programs. We subsequently posted an invitation to participate in a second version of the same survey on the (English-language) listserv of the European Association of Teachers of Academic Writing. When all responses were collected, the N of the WPA-L survey totaled 410, and the N of the EATAW-L survey totaled 65, resulting in an overall sample size of 475.

The survey asked experts 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.” In each case, the survey presented respondents with 10 text boxes in which they could
type words or phrases that they felt suited the two categories. We anticipated that the results would reflect the “threshold writing concepts” deemed important for response to writing in educational settings (see Adler-Kassner and Wardle, 2015). These responses were stemmed using a Porter stemmer, stripped of stopwords (such as “the”), and reshaped into a document-term matrix (DTM), a statistical tool derived from the field of network analysis that allows for meaningful comparisons of word usage across respondents (e.g., Meyer, Hornik, & Feinerer, 2008). The results of additional survey questions, which asked the experts to evaluate their perceptions of the quality of peer and instructor response, were also recorded.

In addition, the survey asked for several demographic responses, including type of academic appointment, primary department or program, overall teaching load, type of institution, age, and gender. Race was not included in the survey, in part because there is an acknowledged lack of diversity in the organizations represented—a problem of continued concern for them—which would have resulted in diminished variation on these variables, rendering statistical tests of racial differences too unreliable to report. Further studies could ameliorate this omission by intentionally oversampling these groups; as our study was designed to provide a representative cross-section of the field and not an oversample, we did not pursue this strategy. Further, we did not expect racial demographics to play a significant role in our expert respondents’ determination of key response terms and concepts, as uptake of concepts from the associated literature on response is expected to have occurred unilaterally across these groups.\(^5\)

**Descriptive Results: Expert Survey**

---

\(^5\)
Before considering the survey responses, it is important to analyze the demographics of the sample. Questions of demographic representativeness are difficult to assess in this case, as the relevant population has not been measured in either census-like or probability samples. However, a description of the sample will help establish a clearer understanding of the basic contours of the group of writing experts.

Based on demographic measures described in Table 1 below, a statistically “average” respondent in our sample is a 50-year old female Associate Professor who teaches composition at a 4-year public university. This typical respondent also claims primary grading responsibility for between 50 and 100 students each semester. There exists substantial variance across many of the measures, however: beyond the central tendencies discussed above, there is evidence that the survey has captured the judgments of a variety of instructors across different career stages, institution types, and demographic profiles. Results of the survey therefore reflect a healthy variety of perspectives from instructors across a number of contexts. Based on the demographic characteristics of the sample and the richness of the content it produced (around 1,700 unique terms), we will consider the most frequently mentioned terms and concepts to represent an expert lexicon for further application to corpora of teacher and student responses.

Next, we consider what the experts thought about response to writing in general. On the basis of Fig. 1, survey respondents generally agreed that response should capture “overall comments on themes, composition, and abstract concepts,” rating this level of commentary as most important to principled response. Fully 94% of respondents ranked such overall
commentary as “most important” relative to the sentence- and paragraph-level commentary, whereas only around 3% of respondents assigned the same importance to sentence-level commentary.

The expert sample clearly mirrors the approaches advocated in the literature on response to student writing. However, Table 2 demonstrates that respondents’ perceptions of the quality of instructor feedback are mixed. The top row of Table 2 shows that the majority (61.8%) of respondents thought that college and university instructors provide response to their students that is of “moderate quality,” while only 12.4% of the sample perceived instructor response to be of “high” or “very high” quality.

Instructors do appear to recognize the potential utility of peer response, however, as 64.1% of those surveyed thought that peer response represents an “extremely” or “moderately” helpful strategy for improving student writing. In general, then, the experts provide an assessment of the state of response that largely echoes the literature on the subject: while the provision of quality instructor response can be hampered by grading pressures or lack of adequate training, peer response is seen as a potentially useful tool to supplement or, at midpoint in the development of a text, even replace instructor response.7

Experts’ Description of Expert and Novice Response
Despite experts’ disagreements in their appraisal of the quality of instructional response, when asked to describe the content of principled and novice response, clear patterns emerge. Fig. 2 provides an initial depiction of the most commonly occurring word stems among survey responses in the form of two word clouds presenting the results of questions asking respondents to provide terms representing principled and novice response to writing.

[insert Fig. 2 about here]

These lists of terms provide preliminary insights about what concepts are important to practitioners in the field, especially when thinking specifically about principled response. Scholars have variously attempted to categorize these considerations on the basis of specific prompt requirements and a priori categorizations, while we do not attempt to create such a typology, it is clear that practitioners are aware of best practices and organize their responses accordingly. A cluster dendrogram, described and presented in the Appendix, shows that many respondents to the survey organized their “principled” terms around overarching concepts like audience (also represented by term like purpose and readership), organization, the development of ideas, coherent structure, and focused support.

Respondents were asked to provide this content-based description of response at the start of the survey, meaning that their responses were unaffected by the potential influence of questions like the ranking item described above in Fig. 1. More complex categorizations aside, the ranked results show clear differences across the most common characterizations of principled and novice response that relate to varying levels of response. Sentence- and paragraph-level
considerations are much more likely to occur in the word cloud representing “novice” response, with terms like “grammar,” “comma,” “spell,” and similar lower-order considerations occupying large areas of the word cloud. Principled response, on the other hand, emphasizes more global and rhetorical considerations such as audience, purpose, focus, organization, support, and clarity.

Fig. 3 demonstrates the relative frequency of these terms in a histogram ranked by the proportion of respondents mentioning each term. One interesting property of this figure is the less robust agreement about novice response. While the average expert survey respondent provided 7.1 unique phrases to describe principled response to writing, experts’ descriptions of novice response contained only 5.0 unique phrases on average. This somewhat sparser response is similarly revealed by the distribution of frequently-mentioned terms in the bottom panel of Fig. 4. It appears that in addition to characterizing novice response as containing mostly sentence- and paragraph-level considerations, experts have a more limited understanding of the response lexicon of novice writers. The descriptive results of our survey confirm a need to better understand the content of peer response to writing for the sake of novice writers and experts alike—and, to the extent that peer response is valued but underutilized in pedagogy, a stronger emphasis on building it into writing curricula.

[insert Fig. 3 about here]

**Study 2: Do Instructors and Peers Practice what the Experts Value?**

Survey respondents provided considerable concept-based insight into their expectations about principled response to student writing. Notably, they have incorporated what appears to be
a strong understanding of the prescriptions of earlier literature into the development of a general lexicon of quality response. Having generated this lexicon, we can apply it to corpus data from student and instructor responses to writing to assess whether the concepts thought to characterize principled response occur in actual response.

Because the nature and content of feedback is expected to vary across different kinds of institutions, curricula, instructors, and assignments, assembling a representative corpus presents some challenges. The present study considers a large corpus of data from a research-extensive institution in the Southeastern United States with a substantial, theoretically informed first-year composition program that adheres to current practices in the field of writing studies. The data were collected as part of an initiative in which instructors utilized software called [redacted for review] to facilitate peer review and to comment on drafts and completed assignments. Data on response to over 1,000 students’ writing assignments in an introductory composition course were collected over the course of two years (2011 – 2013). Overall, the corpus included 49,118 comments from students on their peers’ drafts, and 50,728 comments from instructors in the same courses and on the same papers, totaling 99,846 comments comprising over 3 million unique words. A rubric was utilized by instructors and students to structure their responses to assignments, which included the following criteria for assessment: focus, evidence, organization, style, and format. Students understood that their peer responses would be evaluated based on this rubric. All analyses were coded and executed using the \textit{tm} library in R, a free software environment for statistical computing and graphics.

Both instructor and peer corpora were stemmed (using the same Porter stemmer described above) to facilitate comparisons across the expert survey results and the present data. After removal of stopwords, punctuation, and unnecessary spacing and formatting, the data were
converted into two document-term matrices for analysis. Because of the large size of these matrices, the corpora were truncated through the removal of sparse terms (those terms occurring in less than 1% of comments). The resulting datasets were then analyzed using descriptive frequency analyses, t-test comparisons for the occurrence of individual terms, and correlational analysis.

In previous work using big-data approaches to the analysis of student commentary (e.g., Dixon & Moxley, 2013), fixed lexicons of relevant terms were introduced a priori, and used to conduct analyses of content. Using our expert survey-derived lexicon, we were able to apply more theoretically relevant sets of terms to our analysis. Because the survey asked experts to produce lists of words and phrases they expect to appear in principled and novice response to writing, we developed a lexicon that leveraged the expertise of writing scholars and practitioners who have experience reading and responding to large amounts of writing, and hearing or reading (and sometimes evaluating) student peer response.

**Results: Instructor and Peer Response**

The objectives of this analysis are threefold. First we assess the overall content of instructor and student feedback through simple descriptive analysis of the distributions of frequently occurring terms in the instructor and peer corpora. Second, we evaluate whether these corpora more closely resemble experts’ characterization of “principled” or “novice” response to writing. And finally, we compare students’ and instructors’ use of the most important terms identified in the response lexicon. In doing so, we can provide a detailed assessment of the relative quality and thematic emphases utilized by the two groups.
Fig. 4 presents the most frequently occurring word features in the corpora. The top panel displays a histogram of terms occurring in at least 20% of instructor comments, while the bottom panel provides the corresponding figure for the peer feedback corpus. The results in Fig. 4 provide an initial glimpse at the overall priorities of the commentary, and reveal important similarities across student and instructor feedback. For instance, both students and instructors tended to emphasize a positive tone in their feedback, often using terms like “good” and “great” in their responses. In addition, instructors and students were both quite likely to refer to structural aspects of papers, such as “sentence,” “thesis,” and “paragraph”: a statistically average student in our sample is expected to mention specific sentences in his or her feedback roughly 44% of the time, compared to the average instructor’s likelihood of 41%. It appears that such lower-order, sentential and local-structural concerns are raised by both instructors and students to a relatively high degree.

Differences across the student and instructor content are starker when we consider higher-level concerns, such as the use of sources (mentioned around 36% of the time by instructors, but only around 26% of the time by students). Many examples of such differences emerge when considering Fig. 4, but a more careful analysis across the two corpora can be performed by examining the relative frequency of the terms in the expert lexicon. Below, Fig. 5 presents scatterplots that position the rate of mention of specific words in student and peer comments.
Fig. 5 compares peer and instructor usage of the expert lexicons. The horizontal (x) axis of each panel represents the percentage of instructor comments that included a given term at least once in the instructor corpus. The vertical (y) axis represents the percentage of peer comments that included the same term at least once. The (x,y) coordinates of the plot therefore represent the relative usage of principled key terms across peer and instructor corpora. The solid 45-degree line dividing the plot shows the theoretical intercept at which instructors and peers had equal levels of term usage. This means that any terms falling above the line are used more frequently by instructors than peers, and terms below the line are used more frequently by peers than instructors. Each term is shaded to indicate the rate of mention of the term in the expert survey, meaning that darker terms reflect a greater degree of expert consensus.

Looking first at the left panel of Fig. 5, we can see whether instructors or students were more likely to use terms deemed by experts to constitute aspects of novice or beginner feedback. While instructors were more likely to utilize some terms, such as “need” and “work,” perhaps reflecting a common vehicle for expressing critical evaluations of writing, the most important takeaway from this part of Fig. 5 is the relative proximity of a large number of novice terms around the 45-degree line. Students and instructors do not seem to substantially differ in their likelihood of utilizing content that the experts identified as hallmarks of novice response.

This finding could be interpreted as an indictment of instructors, but looking next at the right panel of Fig. 5, we see evidence of the more frequent use of quality response terms among
instructors. This panel presents a comparison of principled term use across peer and instructor response. Importantly, it appears at first glance that very few of the terms in this panel of the figure fall above the line, meaning that many principled concepts for response are used at least to some extent by instructors. Especially important are terms such as “source,” “specific,” “revise,” “focus,” “argument,” and “develop”: all of these terms appear in instructor response substantially more often than in peer response.

Moving beyond the eye test, Table 3 presents t-test comparisons of the prevalence of the top 10 most important terms in the principled and novice expert lexicons. While many of these differences are statistically significant due to the large N of the samples, 7 out of 10 of the statistical comparisons have a negative sign (indicating greater usage by instructors than by peers). Especially important differences emerge for words like “focus” (δ = 11.9%, p < 0.001) and “develop” (δ = 9.3%, p < 0.001). Whereas students used the word “focus” only 14.1% of the time in their feedback, instructors used the term almost twice as often (a factor of 1.84).

[insert Table 3 about here]

Considering the relative magnitude of the terms across the rows of Table 3, and revisiting the eye test of Fig. 5, a broader pattern of similarity emerges. Across principled terms in the expert lexicon, the Pearson product-moment correlation coefficient $r$ between the rate of mention of students and of instructors is $r = 0.89$, whereas the correlation for novice terms is $r = 0.90$. These very high values indicate considerable similarity in the overall preponderance of terms from the expert lexicon, across instructor and peer response.
Conclusion: A Way Forward for Peer and Instructor Response

The results of the present study have shown that instructors’ feedback often contains the most important components of quality response (as indicated by experts), suggesting that instructor response in a well-informed composition program can exceed the low expectations established by the empirical literature on the subject. Interestingly, peer response in this same instructional context also incorporated many of these same lexical features, though the usage of these features was sparser than instructor feedback to a meaningful degree. Though strongly limited by external validity, the results nevertheless give us important clues about the potential for peer feedback as a useful tool for teaching writing.

The present study is subject to a number of substantial additional limitations. As stressed above, one major area for future research is cross-sectional examination of the content of response, across various demographic and contextual factors known to influence the writing process. In addition, the survey conducted was a convenience sample of writing experts, meaning that we are likely observing a “best case” subset of motivated writing experts who are very involved in the field. In this sense, our study is not a good evaluation of the perceptions of practitioners writ large, despite its utility in providing an expert lexicon. Still further, we also utilize peer and instructor data from just one institutional context, despite the large N of the sample. Further examination of the relationships studied in this paper across institutional context is an important next step for future studies. And finally, the present study presents very simple
descriptive tests of the phenomena in question; future studies should avail themselves of powerful new methods available to researchers in large-scale text analysis, such as discourse network analysis.

Studies of this type are limited by the nature of their a priori measurements of “ideal response” to which empirically-observed response can be compared. By using measures of feedback quality developed by experts, however, we do have some context for comparing instructors’ and students’ responses. These results show clear advantages for instructors in their use of principled feedback—though our students do not lag as far behind as we might have initially expected. A pessimistic approach to these findings might argue that the responses of (novice) students are similar in content to the hurried, superficial, and sometimes unreflective responses of overworked instructors. Despite their more frequent use of principled terms, these instructors’ rates of usage still may not be fully satisfying to some observers.

In contrast, we advocate a more favorable interpretation of the findings which accords with the theory of threshold writing concepts: students may have internalized at least some of the principles of effective feedback through the modeling of their own instructors’ response, through instruction in the core concepts used to talk and think about writing, through preparation for peer review, and through the capacities built into the peer review system. The resulting similarities in content across peer and instructor feedback are therefore a product of the effective implementation of peer feedback. This latter interpretation suggests that peer response can reinforce some of the most important principles of writing development, especially by, strengthen metacognition both for the immediate improvement of a developing text and for the potential transfer of rhetorical and linguistic perspectives and skills to other contexts (as proposed in Yancey, Robertson, and Taczak, 2014). Author (2015), for example, reported on a
study in which students in a first-year composition course created screencasts of their responses to peers’ drafts. Application of a similar lexicon of expert terms based on a smaller survey of experts showed stark differences between two sections of the same course taught by two different instructors. An analysis of the course materials and instructional orientations in these two sections suggested that particular approaches were influencing student responses in the process of peer review. This finding, which should not seem surprising, nevertheless reminds us that students often internalize beliefs and replicate practices modeled by their instructors.

Instruction oriented toward the learning and application of threshold writing concepts is likely to yield the language—and underlying focus—of those concepts in students’ peer responses, while instruction oriented toward the creation of good sentences and the avoidance of error will push students more strongly toward those concerns in their peers’ drafts.

Although further research is needed on the threshold concepts and terms used by both teachers and students in a variety of contexts of response across different genres, results of the present study suggest the potential for stronger faculty-development programs focusing on responding to and evaluating student writing. If instructors expect students to apply principled rhetorical concepts to their peer responses but themselves respond to and evaluate students’ writing in ways inconsistent with those expectations, it is likely that students will either replicate instructional practices or respond in haphazard and inconsistent ways. Modeling expert response and providing direct instruction in how to provide principled peer response may yield more positive results than what many instructors expect and/or experience from peer response.

It is also important to consider the medium in which instructors and students provide response. In this study, students used a digitally-enabled peer review system in which they conveyed their responses in writing, allowing for no discussion or negotiation of options for
revision. This style of peer response differs significantly from the more common method in which students swap drafts, read them beforehand, and then meet face to face in small groups during a class session to share and discuss their responses (see Nystrand and Brandt, 1989). Other technologies, however, may provide different affordances for response. Recently, technological advances in efficient computer-based audiovisual recording have invigorated the scholarship of response to writing (e.g., Author, 2015; Authors, 2016; Author, forthcoming; Jones et al. 2012; Vincelette & Bostic 2013). Screen capture technology allows a student or teacher to create a video recording in which they scroll through a paper onscreen while talking about it, highlighting various sentences or sections and calling attention to various features. This technology has been shown to dramatically improve both the amount and the perceived usefulness of instructors’ commentary (Authors, 2015; Author, forthcoming). Based on the results of the present study, it appears that peer feedback could be amenable to the application of this technology as well. If peer feedback were without higher-order substance, emphasizing only sentence-level grammatical considerations, the affordances of screen capture technology would not substantially improve peer response. However, because peers do appear to provide at least some level of quality content, the affordance of screen capture, with appropriate training and scaffolding, may lead to the stronger presence of those aspects of principled feedback identified by experts.12

Finally, further research is needed to consider other variables that could be related to the language both students and teachers use when commenting on writing. As mentioned, our survey did not capture racial demographics of instructors. Data from the digital peer review system likewise did not provide any information about students’ language backgrounds, race, ethnicity, or other characteristics that would be of potential interest in the study of peer response. Author
and co-author (2010), for example, in a content analysis of instructors’ evaluations of papers written by African American students in their own classes who displayed some AAE linguistics features in their writing, found that instructors frequently misinterpreted, misidentified, mislabeled, and mis-corrected surface elements based on false assumptions and lack of dialect knowledge. Other aspects of race related to the production and evaluation of students’ writing, such as those described in Inoue and Poe (2012), will be important to include in further studies of peer and instructor response (see also Ball et al., 2011).

The results of this study also suggest implications for more deliberate attention to response in faculty-development efforts, both within writing programs and across the curriculum. Response to writing often exists in a “private” domain inside instructors’ classrooms (Author, 1999; 2012). Because high-quality response is crucial in helping students to analyze and improve their own and others’ writing in progress, workshops could ask instructors to analyze and compare the terms and concepts they most often use in their own response against those in the expert survey. Analysis could also focus on the values expressed in the language of response and evaluation, using a process of “dynamic criteria mapping” as described in Broad (2003) and Broad, Adler-Kassner, Alford, and Detweiler (2009). Instructors could also practice working with threshold response concepts, tied to the rhetorical and linguistic focus of their courses, to help students understand, internalize, and apply those concepts in the process of peer response. Across the curriculum, teachers who incorporate discipline-based writing into their courses could be provided with terms and concepts used within the foundational writing program on their campus and encouraged to use them explicitly in their evaluation criteria and any instruction they provide. This alignment of key terms across an entire campus can facilitate students’ transfer of rhetorical and writing-process knowledge as they adjust to often radically different genres, styles,
content, and audiences in their different courses (Author and co-author, 2014; Yancey, Robertson, & Taczak, 2014).

Works Cited


Author, 1989.

Author and co-author, 2010.

Author and co-authors, 2016.

Authors (2015).

Author (1999).

Author (2012).

Author (Forthcoming).

Authors (2016).

Author (2015).


Comer, Denise K., Clark, Charlotte R., & Canelas, Dorian A. (2014). Writing to learn and learning to write across the disciplines: peer-to-peer writing in introductory-level MOOCs. The International Review of Research in Open and Distance Learning, 15(5), 27-82.

Theoretic foundations and new directions. *Review of educational research, 58*(2), 119-149.


Notes

1 Bailey and Garner’s (2010) study is exemplary for the richness of content extracted from instructors on the subject of response to student writing, despite its description of the U.K. context. Studies in the United States strongly corroborate these findings in both public and private contexts: Montgomery and Baker’s (2007) study was conducted at Brigham Young University, while Ferris et al.’s (2013) was conducted in the California state university system.

2 This study was classified as exempt by the Institutional Review Board at [redacted for review], Protocol # [redacted for review].

3 Due to drop-out rates, for some multivariate analyses, N is appreciably smaller.

4 A rough estimate of the active user-base of the WPA-L is 3,500, meaning that the WPA-L response rate exceeded 10%--quite high for the purposes of most expert surveys.

s Despite our inability to directly address the racial dimension of response in this study, racial considerations remain important for future studies of both instructor and peer feedback. This is true especially when we think about the interaction between students and instructors, as we may observe different dynamics across instructor-student dyads of different racial compositions (Kim & Sax, 2009; Lundberg & Schreiner, 2004). Existing research, for example, has pointed to the notion that students may evaluate their peers’ feedback differently on the basis of racial considerations (Pond, et al., 1995). We cannot assess whether instructors or students were of matched or contrasting racial demographics in this study, as our examination of student feedback examines full-sample and not subgroup, patterns—a topic of urgency for future large-scale studies of feedback.
Based upon the mean values of continuous variables and the modal value of categorical variables.

A regression analysis (included in the Appendix) demonstrates that one of the most important predictors of the perceived utility of peer feedback is the number of students graded by the instructor. Contrary to conventional expectations, instructors under higher grading load pressure view peer feedback as a decreasingly useful strategy ($\beta = -0.141, p < 0.01$). It seems that despite the frustrations expressed by instructors in Bailey and Garner’s (2010) study, some instructors appear to be moving away from peer review as a viable supplement or alternative to instructor feedback as their grading loads increase. It is unclear what drives this phenomenon, though we can speculate about how the pressures of extreme grading loads might negatively affect perceptions (or perhaps awareness) of contemporary instructional techniques.

The online Appendix also provides a more exhaustive description of “principled” terms that were mentioned frequently by respondents in the survey.

Comer et al. (2014), for example, use an inductive coding strategy to group comments across 27 unique classifications, including both surface-level and higher-order considerations such as “evidence”, “cohesion”, “provide examples”, and “quotations”.

See [redacted] for a more thorough description of the data generation process.

See [https://cran.r-project.org/web/packages/tm/tm.pdf](https://cran.r-project.org/web/packages/tm/tm.pdf) for more details.

Despite these optimistic conclusions, we must also acknowledge that there exists the possibility that receiving peer feedback in audiovisual modes can heighten “face threat,” a communicative/psychological response to critique that can derail learning objectives. In earlier work,
Author (2015) shows that students do not perceive heightened face threat when receiving audiovisual responses from instructors. Future research is well-positioned to investigate the consequences of affective reactions to feedback, and to delineate such effects from effects on specified learning outcomes.
Fig. 1: Experts’ Perceived Importance of Feedback across Levels of Specificity

The bar chart illustrates the percent of respondents' perceptions of feedback importance at different levels. The levels are categorized into General, Paragraph-Level, and Sentence-Level. The ranking for feedback importance is as follows:

- **Rank 1**: Most Important
- **Rank 2**: Moderately Important
- **Rank 3**: Least Important
Fig. 2: Word Clouds, Expert Description of “Principled” and “Novice” Response to Writing
Fig. 3: Histograms of Most Frequent Terms, Expert Descriptions of Principled and Novice Response to Writing
Fig. 4: Histograms of Most Frequent Terms, University of [redacted for review] Instructor and Student Corpora
Fig. 5: Scatterplots, Instructor vs. Peer Rates of Mention of Novice & Principled Key Terms