Research Article

Computer-Assisted Rhetorical Analysis: Instructional Design and Formative Assessment Using DocuScope

Danielle Wetzel, David Brown, Necia Werner, Suguru Ishizaki, and David Kaufer, *Carnegie Mellon University*

Structured Abstract

- **Background**: Those of us who lead writing programs continue to press toward using writing analytics to better understand how to design, deliver, and assess instruction (Alsop et al., 2019; Reese et al., 2018; Ross & LeGrand, 2017). In particular, corpus-based analytics can offer us rich descriptions of written genres, especially a corpus-based approach that is rhetorical. This approach is especially important in the case of complex tasks of hard-to-assess 21st century communication skills that are often at the center of post-secondary education. In this article, we present an exposition of an ongoing project using DocuScope, a computer-assisted rhetorical analysis tool, for designing and delivering writing instruction within and across courses in the foundational writing curriculum at Carnegie Mellon University. Prompted by a university initiative to champion communication education, like many of our colleagues in higher education, we have been articulating and aligning learning goals, assignments, and rubrics across courses. To facilitate this articulation and alignment, we have used DocuScope as one means of identifying explicit connections between the rhetorical tasks assigned in our foundational courses. This connection between tasks allows us to visualize rhetorical patterns in academic writing and in nonacademic writing, for the purpose of identifying patterns unique to tasks and

patterns that overlap. Both the unique and overlapping patterns can allow us to create a rhetorical discussion around task types, bridging seemingly separate worlds of academic writing and workplace writing—a goal that the contributors share in this special issue. Since DocuScope has been used mainly as a research tool with dictionaries that parse texts with great rhetorical detail and precision, our work has focused mainly upon reconceptualizing the tool for writing pedagogy, with dictionaries aimed to describe categories that build students' textual awareness and rhetorical reasoning. In this manner, we use DocuScope to produce statistical models of tasks across courses in our foundational writing program and then use those models to refine task representation within our instructional design at both a macro and micro level within the classroom. Because DocuScope offers a descriptive and more statistically robust explanation of written genres than other prominent taggers (Brown, 2020; DeJeu & Brown, forthcoming) and because DocuScope prioritizes rhetorical purpose and function rather than lexico-grammatical structures (Ishizaki & Kaufer, 2011), the DocuScope-driven information available to teachers and students is more nuanced and accessible than other types of corpus-driven writing information (Helberg et al., 2018).

- **Research Questions**:
  1. How can the knowledge gained from the computer-assisted rhetorical analysis tool DocuScope shape how we express curricular goals for a writing program?
  2. How can this knowledge not only assist with refining curricular goals but also shape writing instruction that facilitates metacognition and transfer?
- **Conclusions**: We conclude with evidence that what we can visualize broadly, across multiple sections and courses, can also be relevant in the small choices that students make. Students engaging with analytical comparisons and data visualizations experience how genres communicate through fixed and variable language choices. With the DocuScope Classroom tool, programs can make claims about particular curricular goals and align those goals with in-class instruction. We believe this approach facilitates a reconceptualization of assessment as both rhetorical and genre-based, but also as formative for instructional design, informing the vertical integration of writing skills across a curriculum as well as course-level instruction, for both academic and professional writing tasks.

Keywords: computer-assisted rhetorical analysis tool, DocuScope, Workplace English Communication, writing analytics

1.0 Background

To begin, we provide an overview to explain how we interpret dominant assumptions in writing studies and how we see DocuScope Classroom and its identification of language clusters as a means for facilitating student textual awareness and decision-making skills for writing transfer. Often this work is divided between approaches in general writing instruction skills and more specialized writing instruction in programs that focus upon writing in the disciplines, which is an approach to teach students the habits of mind and domain-specific genres that can be used to prepare them for communicating professionally in a variety of career pathways. (For more information about writing in the disciplines, see the WAC Clearinghouse resource page *What is Writing in the Disciplines?*)

In our work, we attempt to bridge these approaches. We extend work on learning transfer within writing studies (e.g., Anson & Moore, 2016; Gere et al., 2018; Hayes et al., 2016; National Research Council, 2012) by building students' textual awareness within a curriculum context that includes a range of academic and workplace genres, as well as within the context of using corpus-based methods within the classroom. Through exposure to patterns of micro-level choices within a variety of writing tasks, students may begin to imagine connections across tasks that previously had been unavailable to them. For such a view of building textual awareness in students, we draw upon Kaufer and Butler's (2000) representational view of writing as a design art, in order to impress upon students the rhetorical effects of their choices. We do emphasize the word "choices," for the writer's decision-making processes must be made explicit if there is to be any strategic, mindful learning transfer.

In this article, we begin with a brief discussion of disciplinary traditions and identification of our research questions. We review how we see our project within the interdisciplinary conversations around writing instruction, genre, transfer, and formative assessment. Like others doing work within education and formative assessment, we value methods that develop habits of mind for rhetorical reasoning and that create spaces for student agency, so that students might be primed with the skills that allow them to adapt to a variety of new writing tasks over a lifetime (Bazerman et al., 2017). We then explain the curricular context of our work by answering each of the research questions and demonstrate the ways our foundational approach incorporates both academic and nonacademic writing tasks. By integrating these tasks in our foundations courses, we avoid the dichotomy of relegating one type of task for foundations and other types for more advanced courses. Instead, we align with the National Research Council (2012) dimensions of expertise in cognitive, interpersonal, and intrapersonal domains, particularly as represented in Table 2.2 of the report on 21st century skills. Finally, we conclude by detailing the ways we have been able to use DocuScope as a tool for representing the assignments across our foundational writing courses and how those representations translate into the particular experiences of individual students making micro-level choices.

2.0 Disciplinary Traditions

All writing instruction and assessment has traditionally been concerned with the following two questions: What kind of assessment information is relevant, and how can this information be made actionable? With an ever-expanding view of what writing is and the role of language, culture, and perspective in writing pedagogy, these questions point us toward formative assessment frameworks that value developing the learner's decision-making. (For an elaboration, see Oliveri, Mislevy, & Slomp, 2021, this issue). As noted by Oliveri, Mislevy, and Slomp in their sociocognitive framework article in this special issue, tensions emerge between traditional forms of assessment (e.g., summative forms of assessment or assessment for reporting) and formative forms of assessment or assessment for learning. The former often focuses on efficiency and reporting of scores, whereas the latter focuses on nuanced learning and enabling opportunities to develop students' agency.

One way to support students as decision-makers is to use learning tools (e.g., rubrics, guided questions, technology) and provide them with opportunities to learn to use linguistic, cognitive, and substantive patterns of language in situated contexts. Such opportunities may prompt students toward understanding to effectively select appropriate genre-specific options rather than favoring a narrow use of tools that aim for measurement only (see Oliveri, Mislevy, & Slomp, 2021, this issue). When we privilege efficiency in the writing classroom, we can abandon the messiness of problem-solving and decision-making. However, there are potential unintended negative consequences associated with this narrower focus, which may lead to academic underpreparation to effectively learn workplace communication English skills that will enable students to be better prepared for future employment or advanced studies. (For an elaboration, see Oliveri, Slomp, Rupp, et al., 2021, this issue). Moreover, David Slomp et al. (2018) remind us that the student's ability to make self-directed decisions is essential for developing writing practices. In fact, this independence should be our ultimate goal.

Traditionally, writing instruction has focused upon structures and genres, but DocuScope requires a view of writing that is representational (Kaufer & Butler, 2000; Kaufer et al., 2004), with rhetorical effects resulting from multiple combinations of linguistic patterns that may not be quickly visible to the writers that choose those patterns nor to the teachers who read their students' writing. For example, writers can notice how many times they use first-person pronouns. This kind of noticing is lexico-grammatical. DocuScope allows writers to see how they are using first-person pronouns for communicating self-disclosure versus for signposting organizational structure. Enabled by DocuScope categories and visualizations, this kind of noticing enables textual awareness and rhetorical reasoning. The assumptions behind using DocuScope dictionaries in the classroom are rooted in design pedagogy (Helberg et al., 2018) and align with current thinking on formative assessment to continue to theorize relevant assessment in terms of disciplinary goals and habits of mind (Andrade et al., 2019; Cizek et al., 2019).

Because writing studies, as a discipline, values language in the particular, problem solving, self-regulation, and learner agency, DocuScope pedagogy highlights the following features: (1) small writing decisions should be treated as important; (2) small writing decisions should be made visible; (3) critical writing decisions should be made visible without sacrificing the complexity of that decision making; (4) computer-assisted analytical tools should provide opportunities for students to develop agency to make data-informed decisions about their writing. Together, these four tenets point to key considerations that have implications related to helping to ensure students are provided with opportunities to learn curricula that include an expanded repertoire of genres and language forms, and include formative assessments that provide students with more meaningful feedback. Such opportunities may enable students to develop metacognition, specifically self-regulation skills, along with other decision-making, based upon the textual information they receive. These opportunities may also boost engagement—Moss et al. (2008) note that when students are provided with highly restrictive approaches to learning, they may disengage and become uninterested.

With these characteristics in mind, while DocuScope is a computer-assisted rhetorical analysis program, it is *not* an evaluative approach to automated, summative scoring. Instead, it is a program that supports student learning by empowering students to make critical inferences using the program's feedback and reflect upon their writing decisions (Helberg et al., 2018). Unlike a program that might draw upon a corpus of student writing as "correct" or "right" in order to evaluate "right" or "wrong" features, DocuScope emerges from writing studies and relies upon human readers and reflection activities for learners. Laura Aull (2018) explains that the former draws upon research from applied linguistics, English for specific purposes, machine learning, and natural language processing. Other approaches are driven by a need to assess writing across multiple contexts, whereas DocuScope as a tool is meant to be used by local readers, writers, and their teachers. Even more specific, DocuScope Classroom has been designed for the classroom context.

Ideally, a corpus-based pedagogy for teaching writing should make visible discursive moves that are typically difficult to see. This pedagogical approach can also help instructors avoid prescriptivism more easily (although never entirely). Instead, instructors can use authentic corpus-based samples (Aull, 2015). Using authentic samples allows instructors to emphasize a normative approach to language rather than a prescriptive one, leading toward a flexible view of language use rather than a fixed, evaluative category. This normative view of language shapes a flexible—and rhetorical—view of genres for the writing classroom, with multiple pathways for resolving rhetorical problems. While most writing teachers would agree that students should be able to reason through multiple pathways for making their compositional choices, we do not necessarily have the time, tools, or feedback mechanisms to give students the assessment information they need to do so robustly.

Undeniably, teachers must make judgments about student progress, and these judgments must align with discipline-specific learning objectives and standards for appropriate language

practices. Assessment information must be relevant to teachers and administrators who must evaluate progress and development. This information must be relevant to students as well, to inform their own goal-setting and decision-making processes about their work—apart from a context of summative assessment (Andrade & Valtcheva, 2009).

To complement other forms of assessment, our goal now persists in developing robust descriptive, formative methods that align with learning objectives, promote student engagement, deepen metacognitive skills, and value disciplinary complexity and habits of mind (Jönsson & Eriksson, 2019). Our goal is to obtain relevant assessment information to enable students to set goals for themselves as writers and to read texts with different lenses in order to make necessary decisions for revision. From our ongoing work within a variety of classrooms, we are seeing that DocuScope provides us with both relevant and actionable information for assessing writing.

3.0 Research Questions

- RQ1: How can the knowledge gained from the computer-assisted rhetorical analysis tool DocuScope shape how we express curricular goals for a writing program?
- RQ2: How can this knowledge not only assist with refining curricular goals but also shape writing instruction that facilitates metacognition and transfer?

These have been ongoing questions for our team over the past six years, balancing the importance of analytical precision and rich data with the very real concerns of the classroom context—which relentlessly demand relevance for learning outcomes. In the classroom, too much information creates noise, becomes overwhelming to parse through, and ultimately stalls learning. On the one hand, DocuScope can give us so much information about a text that the information can overwhelm an untrained reader. A DocuScope analysis can also focus a teacher's eyes away from an immediate learning objective, creating an apparent dissonance between a course's design and its technology-enhanced learning tool. However, we show that if we can use the tool for a bird's-eye view of the curriculum, and if that bird's-eye view coincides with the classroom teacher's view of the curriculum, we might set the stage for a data-informed feedback loop like we have never before seen.

We respond to the first question about curricular goals by using Multi-dimensional analysis (Biber, 1991,1992) to show how DocuScope enables curricular mapping by grouping writing assignments according to clusters of rhetorical effects. When we refer to *mapping*, we mean that we can connect the genres we teach to rhetorical patterns identified through DocuScope. Because we draw upon writing tasks that students produce across academic and professional genres within the foundational space of the university undergraduate experience, we describe how students must travel between the task demands of a fairly wide communication spectrum even before they graduate and take on the demands and rhetorical identities of working professionals.

We approach the second question about the impact of DocuScope for facilitating students' metacognitive awareness through different cases from pilot data collected from classroom

research. These cases are snapshots of course-level interventions with a pedagogical version of DocuScope called DocuScope Classroom. DocuScope Classroom is a learning tool that produces a variety of data-driven feedback to both teachers and students about student writing and compositional decision-making. We provide examples from pedagogical interventions facilitated by DocuScope Classroom to demonstrate how data-informed feedback makes visible particular kinds of information about texts, consequently enabling students to account for their writerly decisions and determine whether they should revise those decisions. In this account, we aim to broaden the field's vision to include pedagogical practice within writing analytics and corpus-based assessment research.

4.0 RQ1: Mapping the Curriculum: Courses, Genres, and Rhetorical Patterns

Multi-dimensional analysis (MDA; Sardinha & Pinto, 2014), enables us to identify and visualize rhetorical patterns at various points and throughout the curriculum, within and across a variety of academic and nonacademic or workplace genres. Before we explain our approach, we describe our curricular context by providing a list of assignment types students encounter. In this section, we also provide a list of the types of language patterns detected by DocuScope dictionaries. Through DocuScope, we have been able to understand the genres in our writing curriculum more deeply because the analyses yield robust information on both fixed and variable rhetorical choices that typically co-occur within and across different genres.

The foundational writing and communication curriculum at Carnegie Mellon emphasizes a variety of genres so that students can gain exposure to a broad spectrum of rhetorical practices. Within the first-year writing courses, students might enroll in a course option that exposes them to introductory nonacademic genres to give them a "view" of communication in workplace contexts. Regardless of course options, we expect students in the first year to learn how to communicate professionally over email and to treat those messages differently from text messages, for instance. As will be elaborated in the following section, we do provide multiple opportunities for academic genres, ranging from the humanities to STEM. And within the 200-level professional and technical writing course, Writing in the Professions, students learn more deeply about professional writing but also consider intersections of professional writing in academic *and* nonacademic contexts as well as how they position themselves as professionals in both academic and nonacademic workplace contexts. Because a DocuScope-based analysis shows that we cannot completely separate academic from nonacademic genres, we decided to reject any notion that professional writing is completely divorced from academic contexts. DocuScope-based analyses enable us to prime a first-year writing student's vision for developing communication skills across the four years, across varying purposes for a lifetime. This curricular design is especially important when we consider writing in workplace settings. As Table 1 illustrates, the curriculum is designed to give students integrated experiences in which academic contexts inform, and are informed by, writing in nonacademic settings. In this way, we are able to provide students some exposure to the kinds of rhetorical choices that occur for

Workplace English Communication (WEC)—explained in the introduction to this special issue and defined as a form of sophisticated discourse in which organizational and disciplinary norms for framing and communicating information are used for a variety of aims (see Oliveri, Slomp, Elliot, et al., 2021, this issue). From a developmental perspective that emphasizes writing knowledge for lifelong learning, writing instruction that only focuses upon classroom and academic genres can limit the long view for developing writers (Bazerman et al., 2018). Providing a variety of genre experiences allows students to engage in comparative thinking, connect similar and dissimilar genre features, and imagine various solutions for rhetorical problems. From both a lifelong and a pragmatic perspective, students need exposure to basic workplace genres and their situations so that they can do work in possible internships that they might win immediately after their freshman year. Many students also solicit opportunities for undergraduate research and must strike up professional communication with faculty that they have never before met. These professional genres are also part of metacurricular activity for those students who wish to participate in civic engagement activities, Greek life, community service projects, and various clubs. Our approach resists some of the literacy theory-based dichotomies attributed to an academic expertise and a professional expertise (Geisler, 1994). Perhaps arguments for abolishing or diminishing mandatory writing classes at some institutions could be avoided if foundational writing programs embraced a wider variety of genres about which students could gain rhetorical knowledge (Wardle, 2009).

4.1 Curricular Context

Within the context of this program, we have collected assignments across a variety of foundational 100- and 200-level courses. Table 1 lists assignments that we have collected from our courses, represented in the genre clusters diagram in Figure 1.

Table 1

Assignments in Courses Across Foundational Writing Course Curriculum

Assignment	Assignment description
Comparative genre analysis	Thesis-driven, analytical essay that compares genre features of either sample texts from different genres or from a set of the same genre type. The goal is to notice fixed and variable structural features as they relate to rhetorical purpose.
Research proposal	An academic, problem/solution/feasibility paper that argues for a research space (using Swales' [1990] "Create a Research Space" [CARS] moves) and that proposes a research plan
Contribution paper	An academic paper that results from the student's research proposal, using either a thesis-driven or Introduction/Methods/Results/Discussion (IMRD) organizational structure (Solacci & Pereira, 2004; Swales, 1990)
Close reading analysis	Thesis-driven, analytical essay that describes and interprets textual observations
Lens analysis	Thesis-driven, analytical essay that uses a theoretical framework to interpret literature or visual art
IMRD report	A data-driven report that both describes and interprets student-generated data analysis
IMRD lit review + proposal	A proposal that synthesizes relevant literature and uses the IMRD organizational form
Cold email	An email message requesting information to a stakeholder recipient who most likely does not know the sender
Memo	A document designed for a busy reader with a clear action item, emphasizing concision, scannable features, and "bottom line up front" organizational structure
Change proposal	A problem/solution/feasibility argument for action that proposes a change to solve a local problem
Cover letter	Part of a job application packet assignment in which students need to write about themselves as workplace professionals

These courses and assignments range from school or classroom kinds of genres ("comparative genre analysis") to more recognizable, portable genres that writers might engage with outside of a classroom context (like a proposal). Note that in describing these assignments, the situations are different enough from each other that students experience different rhetorical purposes and organizational structures. Our guiding principle for this curriculum stems from research in genre theory and pedagogy (Bawarshi & Reiff, 2010): By engaging with this kind of text variety—both academic and nonacademic—students experience comparative genre learning, preparing them for transfer. Additionally, this comparative genre analysis through DocuScope Classroom can

uniquely help students recognize genre-based similarities and differences specifically within and across multiple dimensions of micro-level choices.

4.2 DocuScope and the DocuScope Classroom Tool

The original DocuScope tool with its text-tagging dictionaries was built for researchers analyzing texts (Ishizaki & Kaufer, 2011). The tool is a dictionary-based tagger, meaning that it categorizes words or sequences of words based on a reference list. One well-known example of a computer-readable dictionary comes from Hu and Liu (2004), who placed approximately 6,800 words into two categories—positive (e.g., *fortuitous*) and negative (e.g., *woefully*)—in order to mine the opinions in customer reviews. Another popular dictionary, the Linguistic Inquiry and Word Count (LIWC) dictionary, organizes roughly the same number of words into a variety of categories reflective of its foundations in psychology: affective, social, cognitive, perceptual, and biological processes, as well as drives and orientations. In their partial accounting, Deng et al. (2018) catalogue almost 30 dictionaries, developed from a variety of disciplinary perspectives, including political science, behavioral science, linguistics, and literary studies.

Most human-compiled dictionaries are designed for specific research purposes, varying in size and scope, with the largest dictionaries around 10,000 entries (Deng et al., 2018; Young & Soroka, 2012). Their boundedness is partly a product of the human labor involved in creating them. As Deng et al. (2018) observe, "Developing a special-purpose dictionary is a formidable, iterative, and time-consuming process which could last from months to years" (p. 120). DocuScope itself has been in ongoing development for almost two decades. That unusual development cycle has resulted in an unusually expansive dictionary. The current version tags more than 12 million base patterns and covers more than 98 percent of the total tokens in COCA (Corpus of Contemporary American English).

DocuScope slots those entries into a three-level taxonomy. The version that we used contains 36 categories at the highest level of the dictionary (which DocuScope terms "Clusters"), 3,474 categories at the middle level (called "Dimensions"), and 56,016 categories at the lowest level (called "LATs"). Like any dictionary, its organization reflects its theoretical and disciplinary orientation. DocuScope's orientation is rhetorical. Specifically, it seeks to model the rhetorical effects of writerly decisions—linking the traditionally distinct canons of "invention" and "style" (Kaufer et al., 2004). The dictionary's creator, David Kaufer, explains DocuScope as designed to highlight "the constellation of language choices that provide one or another reading experience to an audience or user" (Kaufer et al., 2004, p. 116). That rhetorical orientation in combination with its size and consequent coverage lend DocuScope robust statistical power in comparisons with other widely used techniques that employ dictionaries, part-of-speech tagging, or both in categorizing linguistic variables (DeJeu & Brown, forthcoming).

For the past seven years, we have been working on a version of DocuScope as a pedagogical tool, DocuScope Classroom, to allow students and teachers to use the hand-curated DocuScope dictionaries to analyze writing in the classroom context. We have also designed a student-

grouping feature that enables teachers to group the most rhetorically different student texts into peer review groups for reflection activities. What follows in this section are some examples of student interactions with the DocuScope Classroom feedback tool and how the students have re-envisioned rhetorical tasks as a result of that feedback on rhetorical patterns. (For a proposed integration of DocuScope into an instance of complex assessments such as those involved with WEC, see Oliveri, Mislevy, & Slomp, 2021, this issue.)

DocuScope Classroom allows teachers to upload exemplars—ranging from expert samples to student samples—and compare their rhetorical features across a variety of different texts. These comparisons can occur through a variety of visualizations, some that emphasize a quantitative view and some that emphasize a qualitative view. An example of a more quantitative view would be a boxplot that shows how a student's patterns (e.g., citation language) compare to patterns used by the rest of the students in the class. (See Figure 4 for an example of a boxplot.) An example of a qualitative view would be a screen that enables a writer to observe those citation patterns, underlined, within the context of the individual paper, or maybe even side-by-side with another paper to compare. (See Figure 3 for a snapshot.) For example, students might be able to see how their favorite writer of a literary analysis paper expressed scholarly humility about their work with hedged confidence language. (In Figure 2, we see one student revising his own confidence language based upon the norms he inferred from expert and other students' samples.) Students can view their language choices in a variety of ways—through a list of words used, through highlighted language in their own writing, through boxplots that plot their favorite author's work in relation to other exemplars, and through scatterplots that compare two features at once.

Using DocuScope Classroom has allowed us to draw upon peer feedback and collaborative learning in the classroom alongside the data-informed feedback. Writing instruction has long valued the significance of multiple perspectives for providing feedback on student writing, and research has shown that students perform better when they receive feedback from multiple reviewers (Patchan et al., 2009). Much of that research on peer review stresses the importance of training students to use rubrics, removing the rubric assessment from evaluation, and using the rubric-informed feedback to set revised goals (Calkins et al., 2019). DocuScope Classroom allows us as teachers to provide rich, focused feedback that neither a single teacher nor multiple peers can possibly give. As the following examples from student work demonstrate, DocuScope Classroom combines multiple dimensions within its visualizations, as well as micro-level views of rhetorical features. In this way, we can help writers see their texts from a profoundly new, data-informed perspective—as if an anthropologist were explaining the culture of their text to them, from an "outsider's" perspective. Students are then forced to reconcile their familiar view of their text with this new, "strange" (Geertz, 1973) way of viewing their text through the lenses of DocuScope Classroom. This process of reconciling various ways of seeing the text leads students to account for their decisions and whether they should keep text or revise it.

Our goal has been to understand the various ways teachers might want to use the data from DocuScope to help their students notice their decisions for key writing objectives. Sometimes those writing objectives align with target genre moves, such as noticing the importance of description in literary analysis. Sometimes those writing objectives seem to align more with complicating students' writing knowledge or expanding that knowledge, such as noticing how citation language can point to a particular relationship with a source text through neutral, authoritative, or controversial stance. But more generally, teachers can use the feedback to build students' awareness of their default language choices. In our student examples that follow later in this article, we see that students confront their small tacit decisions as well as misinformed ones.

4.3 Using DocuScope to Visualize Writing Tasks Across the Curriculum

We can see the above courses and assignments represented in Figure 1. Figure 1 represents a relationship between the assignments and DocuScope categories and is one example of a curricular mapping that explains both pedagogically meaningful and statistically significant variation across several of the tasks in our foundational courses. It is based on the results of multi-dimensional analysis (MDA), which is a statistical technique that is widely used in modeling diverse types of linguistic variation (e.g., Sardinha & Pinto, 2014), including disciplinary and paper-type variation in academic writing (Grey, 2013; Hardy & Römer, 2013). The figure plots the mean dimension scores for paper assignments in four different first-year writing courses (101, 106, 107, and 108) and a professional writing course (270). See Table 1 for a listing and a description of those assignments. The scores are derived from the DocuScope categories that have positive (> 0.35) and negative (< -0.35) factor loadings along a given dimension, following conventional MDA thresholds (Biber, 1991, 1992). Along the first dimension, for example, First Person, Positive, Metadiscourse Interactive, and Future have high positive loadings, while Information Exposition, Reasoning, Academic Writing Moves, Citation Authority, and Confidence Hedged have high negative loadings. (For descriptions of the categories, see Table 2.)

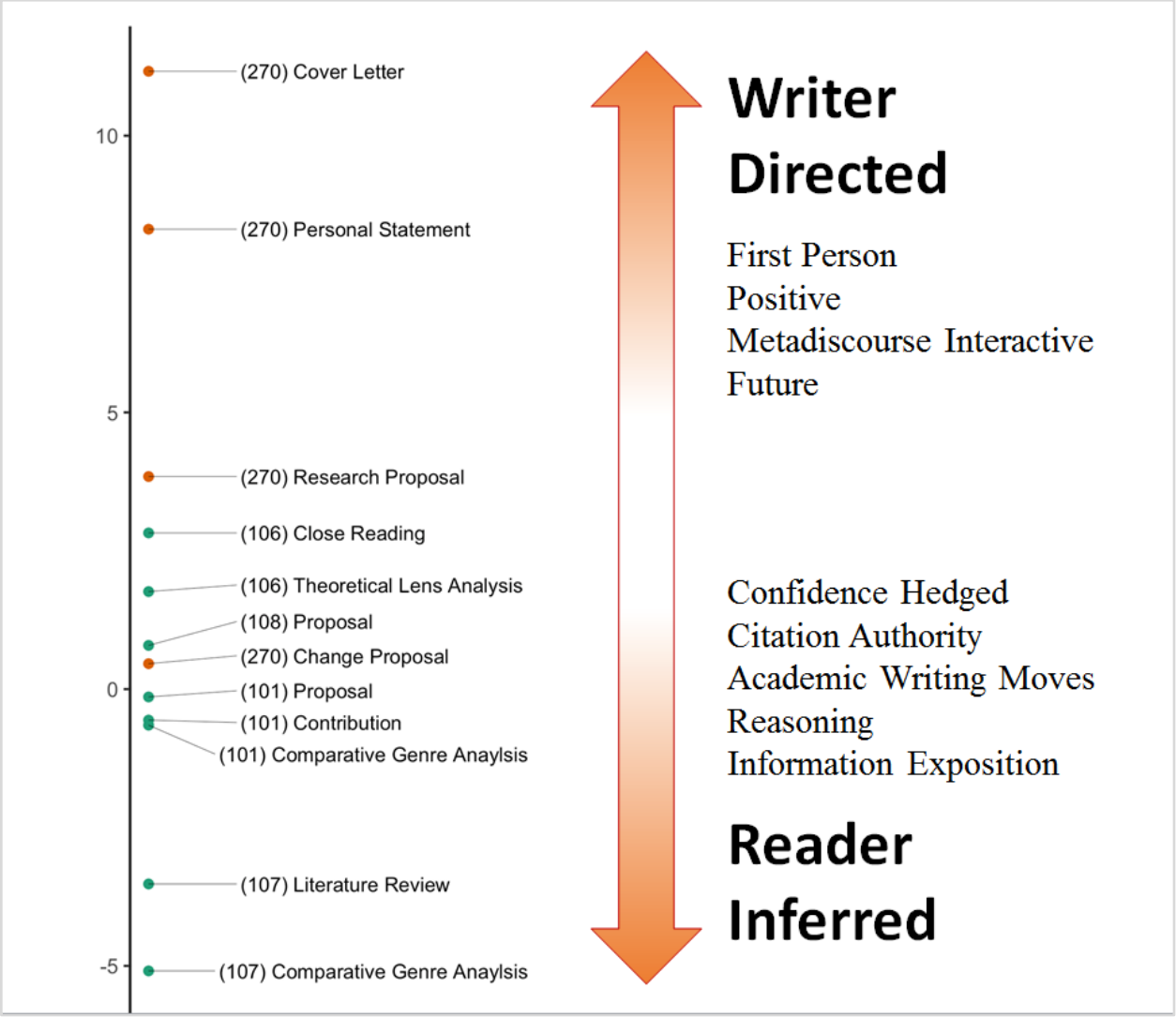
The position along a dimension is determined by summing the low and high loadings for each paper, subtracting the low from the high loadings to arrive at a "score," and calculating the mean score for each writing task. At the positive end of the cline along the first dimension, there is a higher concentration of positive, Writer Directed features and fewer negative, Reader Inferred features, and vice versa. Based on a one-way analysis of variance (ANOVA) with over 1,000 papers, a statistically significant effect was found for the Writer Directed and Reader Inferred dimension at $p < 0.0001$. $F(11, 1430) = 428.10$, $p < 0.0001$, accounting for 76.53% of the variance. This result, informed by an automated analysis of over 1,000 papers, would have been impossible to glean from human serial reading. Moreover, because of DocuScope's focus, that description allows us to see what is rhetorically similar and different across tasks, and the

types of linguistic modifications students need to make to move between genres and rhetorical structures (Ringler et al., 2018).

In Figure 1, we see groupings around three types of genres—informational, argument, and professional documents about self—across writing tasks in the writing courses. The first grouping at the bottom represents the most informationally dense academic writing along the spectrum—akin to the explanatory genre family that Laura Aull (2019) describes in her work on stance in undergraduate-genre families. The more a genre moves toward a technical report with an IMRD organizational style (defined above), the more the features point to compositional choices that position a reader to inference discourse community and meaning through hedged confidence, authoritative citation, academic writing moves (similar to those described by Swales, 1990), and information exposition. At the middle of this spectrum, we see a cluster resembling what Aull (2019) calls an argument genre "family," distinctive from the explanatory-genre family. In the corpus of papers outlined in Figure 1, these genres are academic arguments, humanistic literary critiques and research-based argument papers, research and change proposals driven by problems, and arguments for feasible solutions.

Figure 1

Mapping Showing Clusters of Genre Moves Ranging Across the Writing Tasks from Five Different Writing Courses



Note. This mapping shows clusters of genre moves ranging across the writing tasks from five different writing courses (See Table 1, 76-10X and 76-27X), identifying combinations for refining curricular goals and instructional materials to facilitate vertical integration of rhetorical skills for learning outcomes. [One-way ANOVA: $F(11, 1430) = 428.1$, $p < 0.0001$, $R^2 = 76.53\%$]

The other end of the spectrum shows rhetorical choices that point to the genres of a professional cover letter and statement. These are labeled "Writer Directed," higher in first-person pronouns, positive language, interactive types of metadiscourse, and future-oriented language. In Table 2, we present a listing of DocuScope categories. As stated earlier, the

combination of genres prompts students to engage with language combinations and rhetorical patterns they might not otherwise prioritize in traditionally humanistic and STEM academic genres.

Table 2

DocuScope Category Definitions and Selected Examples

Category (Cluster)	Description	Examples
Academic terms	Abstract, rare, specialized, or disciplinary-specific terms that are indicative of informationally dense writing	*market price*, *storage capacity*, *regulatory*, *distribution*
Academic writing moves	Phrases and terms that indicate academic writing moves, which are common in research genres and are derived from the work of Swales (1981/2011) and Cotos et al. (2015, 2017)	*in the first section*, *the problem is that*, *payment methodology*, *point of contention*
Character	References multiple dimensions of a character or human being as a social agent, both individual and collective	*Pauline*, *her*, *personnel*, *representatives*
Citation	Language that indicates the attribution of information to, or citation of, another source.	*according to*, *is proposing that*, *quotes from*
Citation authorized	Referencing the citation of another source that is represented as true and not arguable	*confirm that*, *provide evidence*, *common sense*
Citation hedged	Referencing the citation of another source that is presented as arguable	*suggest that*, *just one opinion*
Confidence hedged	Referencing language that presents a claim as uncertain	*tends to get*, *maybe*, *it seems that*
Confidence high	Referencing language that presents a claim with certainty	*most likely*, *ensure that*, *know that*, *obviously*
Confidence low	Referencing language that presents a claim as extremely unlikely	*unlikely*, *out of the question*, *impossible*
Contingent	Referencing contingency, typically contingency in the world, rather than contingency in one's knowledge	*subject to*, *if possible*, *just in case*, *hypothetically*

Category (Cluster)	Description	Examples
Description	Language that evokes sights, sounds, smells, touches and tastes, as well as scenes and objects	*stay quiet*, *gas-fired*, *solar panels*, *soft*, *on my desk*
Facilitate	Language that enables or directs one through specific tasks and actions	*let me*, *worth a try*, *I would suggest*
First person	This cluster captures first person.	*I*, *as soon as I*, *we have been*
Force stressed	Language that is forceful and stressed, often using emphatics, comparative forms, or superlative forms	*really good*, *the sooner the better*, *necessary*
Future	Referencing future actions, states, or desires	*will be*, *hope to*, *expected changes*
Information change	Referencing changes of information, particularly changes that are more neutral	*changes*, *revised*, *growth*, *modification to*
Information change negative	Referencing negative change	*going downhill*, *slow erosion*, *get worse*
Information change positive	Referencing positive change	*improving*, *accrued interest*, *boost morale*
Information exposition	Information in the form of expository devices, or language that describes or explains, frequently in regards to quantities and comparisons	*final amount*, *several, three*, *compare*, *80%*
Information place	Language designating places	*the city*, *surrounding areas*, *Houston*, *home*
Information report verbs	Informational verbs and verb phrases of reporting	*report*, *posted*, *release*, *point out*
Information states	Referencing information states, or states of being	*is*, *are*, *existing*, *been*
Information topics	Referencing topics, usually nominal subjects or objects, that indicate the "aboutness" of a text	*time*, *money*, *stock price*, *phone interview*

Category (Cluster)	Description	Examples
Inquiry	Referencing inquiry, or language that points to some kind of inquiry or investigation	*find out*, *let me know if you have any questions*, *wondering if*
Interactive	Addresses from the author to the reader or from persons in the text to other persons. The address comes in the language of everyday conversation, colloquy, exchange, questions, attention-getters, feedback, interactive genre markers, and the use of the second person.	*can you*, *thank you for*, *please see*, *sounds good to me*
Metadiscourse cohesive	The use of words to build cohesive markers that help the reader navigate the text and signal linkages in the text, which are often additive or contrastive	*or*, *but*, *also*, *on the other hand*, *notwithstanding*, *that being said*
Metadiscourse interactive	The use of words to build cohesive markers that interact with the reader	*I agree*, *let's talk*, *by the way*
Narrative	Language that involves people, description, and events extending in time	*today*, *tomorrow*, *during the*, *this weekend*
Negative	Referencing dimensions of negativity, including negative acts, emotions, relations, and values	*does not*, *sorry fo*r, *problems*, *confusion*
Positive	Referencing dimensions of positivity, including actions, emotions, relations, and values	*thanks*, *approval*, *agreement*, *looks good*
Public terms	Referencing public terms, concepts from public language, media, the language of authority, institutions, and responsibility	*discussion*, *amendment*, *corporation*, *authority*, *settlement*
Reasoning	Language that has a reasoning focus, supporting inferences about cause, consequence, generalization, concession, and linear inference either from premise to conclusion or conclusion to premise	*because*, *therefore*, *analysis*, *even if*, *as a result*, *indicating that*
Responsibility	Referencing the language of responsibility	*supposed to*, *requirements*, *obligations*

Category (Cluster)	Description	Examples
Strategic	This dimension is active when the text structures strategies activism, advantage-seeking, game-playing cognition, plans, and goal-seeking.	*plan*, *trying to*, *strategy*, *decision*, *coordinate*, *look at the*
Syntactic complexity	The features in this category are often what are called "function words," like determiners and prepositions.	*the*, *to*, *for*, *in*, *a lot of*
Uncertainty	References uncertainty, when confidence levels are unknown	*kind of*, *I have no idea*, *for some reason*
Updates	References updates that anticipate someone searching for information and receiving it	*already*, *a new*, *now that*, *here are some*

As advocates for our writing programs, we have found that this mapping in Figure 1 to the discourse practices identified in Table 2 enables us to point to a variety of skills that our students are producing across tasks. We can project a developmental trajectory for writing that captures disciplinarity (note the STEM-oriented groupings of 107 papers from Writing About Data) and also discursive aim (note the evidence-based, argument groupings from essays to proposals and then the genres oriented toward professional self-fashioning, for which students must learn to represent themselves as professionals and communicate that representation in a compelling way).

However, while we acknowledge that this mapping allows us to see clear distinctions between tasks for instructional design purposes, we are keenly aware that such efforts are a bird's-eye view that may not have great relevance in the classroom experience. In other words, can individual writers making particular decisions for their texts find any help in the kinds of information that DocuScope assesses across the program? This question, of course, is the subject of future research. Our preliminary findings suggest that we are able to see evidence of linkages between students' interactions with DocuScope and DocuScope's analytic power based on the results of multidimensional analyses. Next, for our second overarching research question, we see how individual students interact with these categories within their particular texts and show some awareness of those categories in their writing. These linkages confirm that some of the statistical patterns on a large scale can be useful at the smaller scale through the use of a pedagogical tool, DocuScope Classroom.

5.0 RQ2: Translating DocuScope Findings into Classroom Interventions

Our second research question prompts us to connect the classroom experience with the knowledge gained from a multidimensional analysis of DocuScope categories within student writing. Throughout our iterative work piloting the DocuScope Classroom tools, we have experienced the typical frustrations that occur with developing new technological tools. But we should also note an additional challenge that we associate with teaching writing with tools that do not give students scores or definitive answers but rather prime students to make inferences. As discussed earlier, students have been primed to view literacy through testing, in terms of right and wrong answers (Hillocks, 2002; Trachsel, 1992), or students have been socialized to see writing knowledge as completely subjective and dependent upon an individual's perspective. Yet despite these challenges, we can point to stories that encourage us that we are on the right track even if descriptive, formative feedback feels unfamiliar at first to students. We share some of our pilot data stories collected from the context of classroom-based research. (This study was approved by the Institutional Review Board [IRB] of Carnegie Mellon University.) We give three examples of students' engagement with course-level interventions informed by DocuScope Classroom visualizations, connecting some of the strategies present in the Figure 1 mapping. Through different pedagogical interventions facilitated by DocuScope Classroom, we can see how data-informed feedback heightens textual awareness and rhetorical reasoning skills, enabling students to account for their writerly decisions and determine whether other decisions might be better.

The first two examples point to students' revision processes enabled through DocuScope Classroom data. The third shows a student's annotations in a Google Doc after reading a data-driven report produced by DocuScope Classroom, troubleshooting composition problems with the DocuScope-enabled language. The first and third examples are from students enrolled in our 100-level course, Writing about Literature, Art, and Culture. In our second example, a student enrolled in a 200-level course, Writing in the Professions, grapples with choices that may signal too much negative language for a cover letter, according to a DocuScope Classroom report, and ultimately revises due to concerns that a reader may be turned off by too much negative language.

5.1 DocuScope Classroom Information Prompts Students to Revise Texts

Our first two student examples, Jared and Steve, both show how DocuScope Classroom feedback assisted their revision processes. The methods that the teachers used to prompt their reflection and revision were different, and the course contexts were different as well. These differences point to the rich possibilities for implementing the tool.

5.1.1. Example 1: Jared Wrestles with High Confidence Language for Academic Writing

When they enter the first-year writing classroom, students bring their prior assumptions about what makes a strong argument in academic discourse. Some students have participated in oral

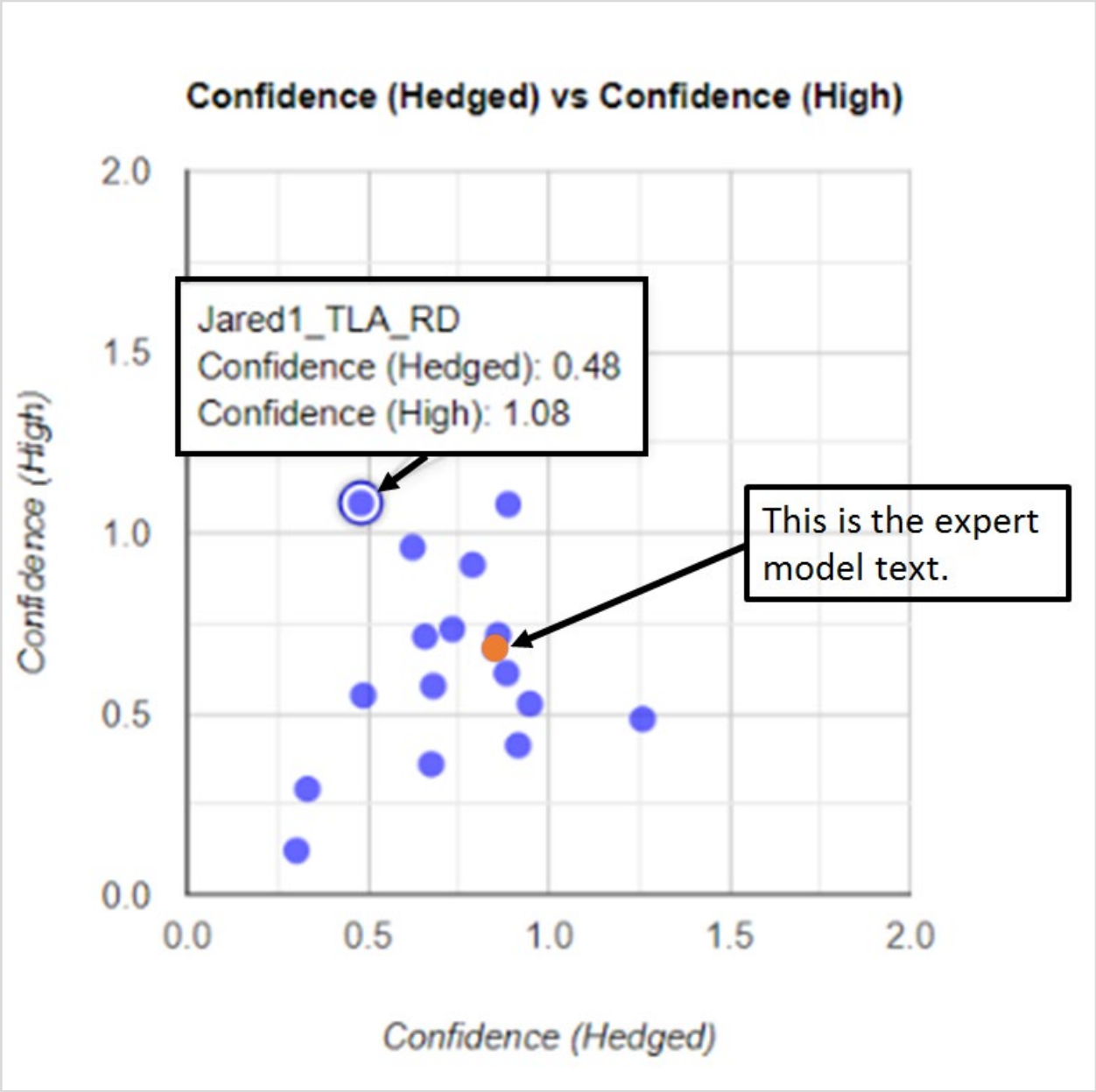
debate in high school, they have written position papers for competitions, and they have observed extreme political discourse in our media (in the United States). The language that writers use to construct credible, persuasive arguments can vary across contexts and genres—but oftentimes students have had only one encounter with the concept of argument.

One of the learning objectives in first-year writing courses is that students will be able to write claims with language that fits the weight of their evidence and explanation. Persuasive writing for many students is the equivalent of making claims that use language that "sounds persuasive," apart from any discourse community expectations. Brown and Aull's (2017) findings on emphatic generality versus elaborated specificity point to some of the routine ways that students write academic arguments for Advanced Placement (AP) tests, with lower-scoring essays using boosters to signal higher, more confident, levels of certainty for claims. These kinds of boosting claims can occur within first-year writing texts. Often these claims do not represent the nuances characteristic of academic writing.

Jared, a dedicated student in a first-year writing class called Writing about Literature, Art and Culture, received feedback on his second Lens Analysis draft from DocuScope Classroom. The instructional approach within this course included opportunities for early feedback on small pieces of the project, and these opportunities employed a variety of different types of feedback. Some of this feedback included peer/pair/share feedback on developing ideas in body paragraphs, instructor comments on preliminary thesis and outlines, and self-assessments utilizing methods like a reverse outline. After students had produced two versions of the paper, their teacher gave them DocuScope Classroom reports of their language choices and paired the students for a reflection activity with guiding questions. However, this reflection activity alone was insufficient to catch Jared's attention. Instead, Jared was moved to notice his language choices (qualitative view) after viewing the scatterplot view (quantitative view) with the instructor. When viewing the student drafts along a scatterplot, Jared compared his work with his peers' written work and also with the model texts they had read previously in the course. These model texts included model student papers as well as expert samples. Figure 2 shows a scatterplot view in DocuScope Classroom, meant to help students focus upon two features that teachers might often pair together to provide contrast and perspective on language choices. For this example, the instructor wanted students to consider their decisions for the strength of their language (similar to hedges and boosters) for representing the certainty of their claims. This view enabled Jared to observe his text as higher in "Confidence High" language than any other text in the group.

Figure 2

Scatterplot View in DocuScope Classroom



Note. Scatterplots allow students to see how their language choices fit in and stand out within a group, prompting opportunities to account for writing decisions. Jared, a first-year writing student, learned through a scatterplot visualization that he used more *high confidence* language than anyone else in the class, including expert models.

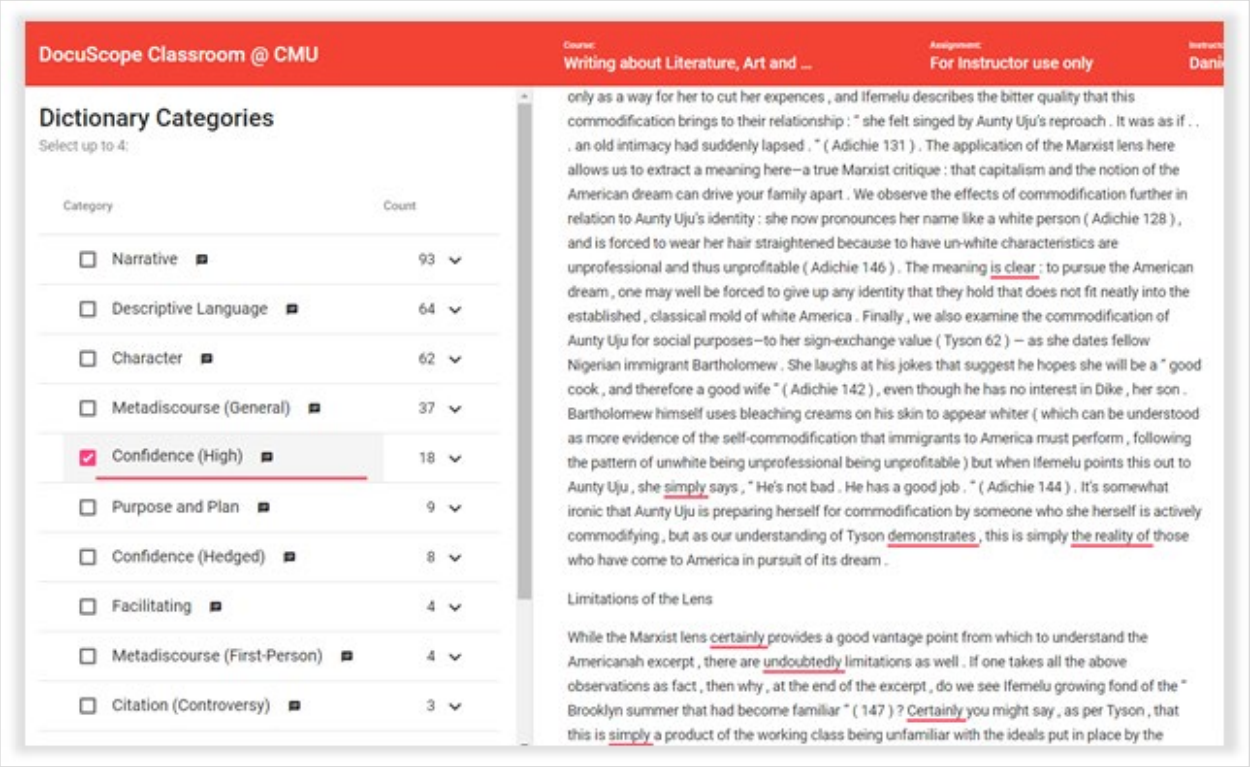
Earlier in Figure 1, we can see that "Confidence Hedged" is a category that is distinctive to a category of assignments in the multidimensional analysis, particularly for academic texts. DocuScope Classroom enabled Jared to pay closer attention to how certain his language sounds

for an academic paper, and he was able to compare his choices with other writers' choices, both expert and novice samples. The point of course was not that all writers should be exactly the same but rather that writers must consider how their choices point to the kind of company they keep.

Because scatterplots are not readers and do not provide us with a description of text quality or the context of the writing in which patterns occur, Jared's investigation was just beginning. After Jared noticed that he wrote with higher levels of certainty than other students in the class, he read his text through the lens of questioning whether he truly wanted to use all of the language choices that expressed high confidence. In Figure 3, another view of Jared's text in DocuScope Classroom allowed him to read "Confidence High" language highlighted in his text alone. This language included patterns like *is clear*, *simply*, *demonstrates*, *certainly*, and *undoubtedly*.

Figure 3

DocuScope Classroom Language Patterns View



Note. A necessary step in any DocuScope Classroom activity is that students must read language patterns within the context of writing. Jared read his high-confidence language within his writing to determine whether his levels of certainty were indeed what he wanted to communicate to his reader.

Ultimately, Jared used the comparative information from the scatterplot visualization, along with a careful reading of his report that identified his "Confidence High" language choices, to revise his paper. When he asked to see another visualization of his text in relation to his former

one, DocuScope Classroom showed Jared how his text compared to his previous draft. Jared did not remove all of his "Confidence High" language. Instead, he reviewed and accounted for his choices to ensure that they were the best ones he could make for communicating the argument that resulted from his analysis. In this way, the automated feedback from DocuScope Classroom still kept the writer firmly in the driver's seat.

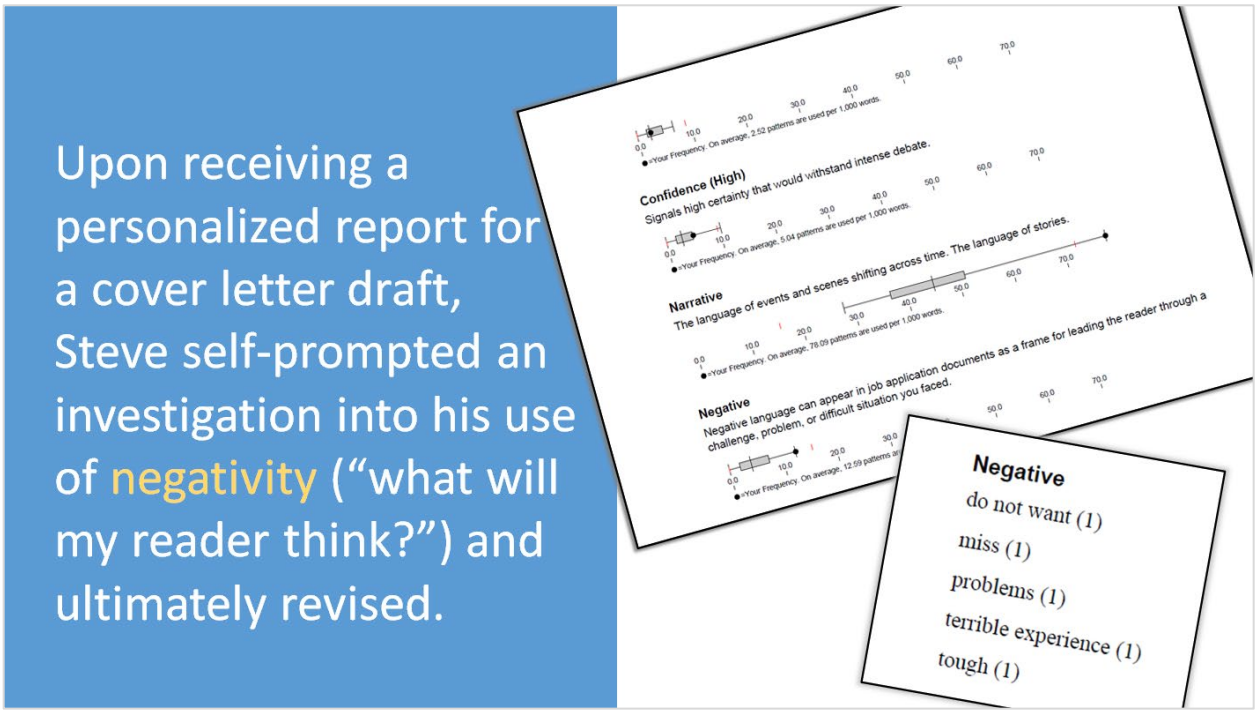
5.1.2 Example 2: Steve Wrestles with Negative Language in Writing in the Professions

In a professional writing course, some of the most important feedback from DocuScope Classroom is the feedback that points to what we call "small choices" that combine with others in a particular genre. The use of positive language is one of those small choices in a cover letter that is distinctive in the groupings in the mapping represented in Figure 1. For the cover letter assignment, students write for audiences they have never met and can only interact with readers through a first impression. First impressions, as social interactions, can be encouraging and uplifting—prompting a reader to want to know more about the writer. But first impressions can also fall flat, creating a disconnect, prompting a reader to desire an ending rather than a beginning. Negative language in a cover letter can function in a manner that disconnects readers from writers and hastens the end of a social interaction. The role of negative language can be powerful when describing problems, but not necessarily when writing about the self or professional experiences within the context of a cover letter. For many students who are just beginning to see themselves as professional people, the compositional challenges of simply selecting what to say about themselves are so great that they cannot hold in their working memory other considerations due to audience expectations and discourse community (Kellogg, 2008).

In Figure 4, we see that Steve's teacher had provided students with DocuScope Classroom reports of their writing, and these reports included a boxplot view of the students' texts that allowed them to compare with each other. We can see from the description of *negative language* that within the context of the Writing in the Professions class, students are encouraged to think of uses of negative language to "lead the reader through a challenge, problem, or difficult situation." In other words, negative language is not bad language to use in a cover letter but rather purposeful language. After noticing his negative language choices and reading them within the context of his letter, Steve was not convinced these were the best choices he could make.

Figure 4

Boxplot View of Texts in DocuScope Classroom



Note. Through boxplots that prompt comparison with other writers' language choices, along with a language listing identified as *negative*, DocuScope Classroom facilitates an opportunity for a writer to examine small choices that do not signal a positive, forward looking cover letter.

Like the scatterplot that showed Jared how his choices compared with other writers' choices, the boxplots in Figure 4 provided Steve a quick glance at how his negative language choices were so high that they were statistical outliers compared to the rest of the class. Because the boxplots only give a normative view of language choices, it is important to examine the language—Steve discussed problems, he referred to losses with the word *miss*, and he used *terrible experience* to narrate an event. With DocuScope Classroom feedback, Steve re-examined his choices for framing a part of his cover letter through this negative language. He ultimately revised his cover letter to contain fewer instances of negative language that pointed to attitude or state of mind.

Both Jared and Steve responded to the feedback and visualizations from DocuScope Classroom in a reflective, thoughtful manner. We would like to note that this feedback did not include scores of any kind. Students were not given grades based upon DocuScope Classroom information, nor were they told some profiles were better than others. Instead, the students were asked to consider their choices and account for them.

5.1.3 Example 3: Jaden Annotates Her Google Doc with DocuScope Classroom Concepts

We wanted to include the example from our third student, Jaden, because it surprised us. We did not anticipate how students would produce annotations in their own Google Docs—and Jaden's annotations were not unique to only her. Other students (outside of her peer review group) also produced similar annotations in their writing. For this activity, the students' writing was not compared with expert writing (as was the case with Jared's text). Instead, they identified for each other which parts of their texts were not clear, and then students targeted DocuScope Classroom language that might help them clarify particular areas in their texts.

Like Jared in Example 1, Jaden was a first-year writing student struggling with a Lens Analysis draft, albeit in a different section of the course. Jaden and the other students in the course had encountered DocuScope Classroom concepts tailored for their course, with naming conventions somewhat different from Jared's—yet the instructional approach was similar. These concepts were presented to students in the following ways:

- **Reader-directed metadiscourse** includes language that orients readers to the world of the text. This language helps readers understand the writer's structure, purpose, and logic. Metadiscourse shows readers where they have been and where they are headed, and helps them experience the writer as a "tour guide" to the experience of the text.
- **Purpose & Plan** includes language that signals the purpose, need, and direction of the argument.
- **Facilitating** includes language that communicates how something or someone serves as a help, a guide, or an encouragement. For our class, facilitating language can help us show readers the benefits and affordances of interpreting a text with a particular lens or framework.
- **High Confidence** includes language that signals high certainty that would withstand intense debate.
- **Hedged Confidence** includes language that signals moderate certainty and openness to other points of view.

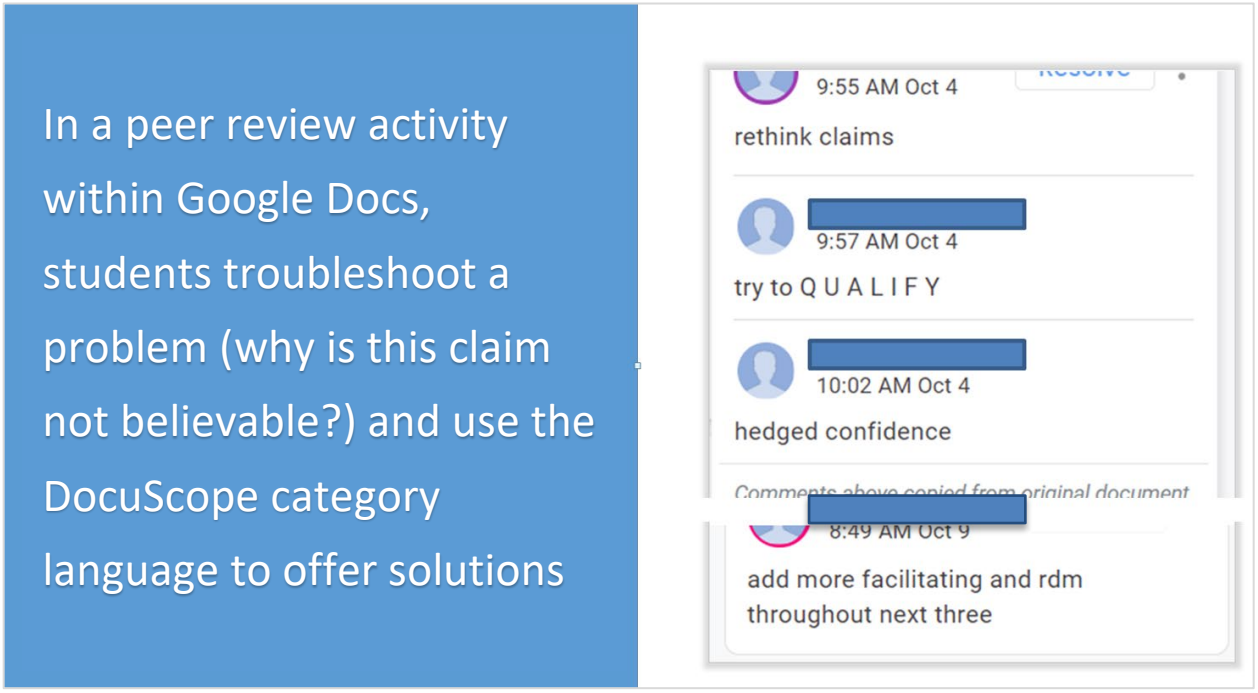
Students received multiple types of feedback for this task, including preliminary feedback on pieces of their Lens Analysis drafts, peer feedback based upon guided questions and a reverse outline activity, and then DocuScope Classroom reports with guided questions to reflect on their decisions. After reviewing peer feedback together, the students were given class time to annotate their drafts with notes that would help them move toward revision.

The student annotations in Figure 5 represent one writer's process for troubleshooting claims in her thesis-driven, interpretive essay. Note that Jaden wrote about the strength of her claims the most, which is not surprising given the amount of instructional time allotted in the class for focusing students on interpretive claims for academic arguments. Jaden's focus upon hedged confidence mirrors Jared's earlier concerns regarding the amount of high-confidence language

present in his writing and the cluster mapping in Figure 1. We also see Jaden's processing of what is the function of using hedged language—to "qualify." The mixing of DocuScope Classroom categories along with an imperative statement to self, "Try to qualify," points to a fairly strong understanding of the concept. And four days later, the student returned to her text to point to a need for more "rdm" (reader-directed metadiscourse) and "facilitating." This focus pointed to the student's awareness of using language to help readers navigate the high-priority claims and takeaways of her analysis.

Figure 5

Student Annotations Using DocuScope Classroom Concepts



Note. After hearing a lecture on argument claims for workshopping drafts and after receiving a DocuScope Classroom report about her work, Jaden annotated her draft with concepts from DocuScope Classroom, offering a glimpse of uptake within the class.

In Jaden's first annotations, she focuses upon the content of her argument and the certainty of her claims. She used the concept of *hedged confidence* to help her approach rewriting. Her second note, written five days later, included DocuScope Classroom language related to concepts that help writers guide readers through their text to navigate structure and to notice high-priority statements. We realize that Jaden's notes do not represent hard evidence for learning, although she was not the only student from her class who wrote those kinds of notes in Google Docs. We find it curious to see that within Jaden's own reflections about her writing, she used the concepts from DocuScope Classroom to guide her revision process. We believe that Jaden's notes

represent how DocuScope Classroom feedback, when aligned with instructional priorities and genre expectations, can become a metacognitive tool for student-driven, formative assessment in the kind of goal-directed ways that we hope point toward independent practice.

These three examples combined point to different use cases "in the small" that encourage us to maintain the claim that descriptive, rhetorically-based statistical analyses can align with classroom instruction. DocuScope Classroom provided descriptive feedback that enabled students to see their texts differently, compare and evaluate their decisions with expert and peer texts, and subsequently account for their decisions. With these examples, we can see how a DocuScope Classroom assessment provides formative feedback that is both relevant and actionable.

6.0 Conclusion

From the start, we posed broad questions about whether descriptive, corpus-based rhetorical analysis can be useful to shape macro-level writing curriculum goals that will also translate into classroom instruction and meaningful learning experiences that prime students for transfer. We answered the first question about how DocuScope might inform curricular goals by presenting a multidimensional analysis of rhetorical features from a corpus of student writing produced across a variety of genre tasks in a foundational writing program. The visualization in Figure 1 shows how a bird's-eye view of a curriculum, because of the clear groupings around target genres, allows us to make claims about rhetorical benchmarks that students achieve. Additionally, the bird's-eye view enables a relational view of rhetorical patterns along a range of genres. This relational view combines academic and nonacademic genres within a discursive world that explains genre through comparing language choices and rhetorical purposes, prompting us to articulate connections and highlight differences among the number of linguistic options on the rhetorical design palette. We can also compare academic genres and workplace genres, helping us to explain why students can leave the university and then encounter writing tasks in the workplace with difficulty, a concern that many researchers have demonstrated since learning transfer has emerged in scholarship from general writing instruction and writing in the disciplines (e.g., Beaufort, 2007; Conrad, 2017).

For our second question about how a macro-level view of written genres might inform the individual student experience and facilitate metacognition for transfer of writing knowledge, we offered three student examples as individualized responses to the DocuScope Classroom feedback. The genre information from DocuScope Classroom suggests ways of refocusing attention to the text. Ultimately, by highlighting how rhetorical features connect to course tasks, teachers focused their students' attention in particular ways, just as we saw with the example of Steve above when he reflected upon his use of negative language in his cover letter. The bird's-eye view of the DocuScope-informed analysis enables an awareness of features that can refine the kinds of genre-oriented feedback students receive on their writing. In the case of Jared, his default use of high-confidence language made him question how he represented the certainty of

his claims. In the case of Jaden, focusing upon citation and high confidence, as well as other features, enabled her to link language choices to particular genre aims such as writing clear claims and explaining the affordances of interpreting a literary text with a theoretical lens. DocuScope Classroom, in the student examples, provided students with targeted information about relevant-genre features and also allowed students to compare their choices with other writers' choices.

We believe we achieve greater coherence and integrity for a writing curriculum if we can articulate a data-informed range of rhetorical choices within a sequenced curricular experience. Our work includes designing and using tools that enable teachers and students to access descriptive feedback about those specific rhetorical choices. This feedback is useful only if it is formative, prompting students toward greater textual awareness to employ rhetorical reasoning about options that make sense for readers. The student examples above—from Jared, Steve, and Jaden—all point to various ways that a data-informed approach, through DocuScope and DocuScope Classroom, might shape writing instruction by creating opportunities for students to view models, write, compare, rewrite, and reflect together about their choices.

DocuScope Classroom provides ways for instructors to design activities that promote reflection and revision in students' writing processes. As we approach another year of instruction with DocuScope Classroom, we look forward to new features in the platform that will enhance student and teacher options within our course management system, Canvas. We also look forward to collecting more data about not only how students are using the tool but also how teachers are using the tool. We do have some preliminary evidence that our multilingual students who are nonnative English speakers benefit uniquely from the tool. Moreover, we expect to pursue projects on how DocuScope Classroom can teach underserved, linguistically diverse students to see writing as a process of normative decision-making rather than as a prescriptive set of rules. Moving forward, along with continuing to build an understanding of how teachers align their instruction with the tool, we hope to learn more about how less-experienced teachers approach and use DocuScope Classroom. Ultimately, our ongoing projects must include developing a program for scaffolding teachers, with varying degrees of expertise, to engage independently with the tool.

We conclude with these statements about the affordances of using data-informed tools that combine language and purpose, such as DocuScope (the research tool) and DocuScope Classroom (the pedagogical tool). Through multivariate analysis enabled by DocuScope, we can produce statistical models of task representation that enable vertical integration of skills across a curriculum, and we can use statistical models of tasks to inform our writing pedagogy. Moreover, we can produce data visualizations on how genres communicate through fixed and variable language choices and use those visualizations in the classroom to assist students' task representation and writing processes. From a bird's-eye view, we can bridge the gap between university and workplace writing by mapping genre features according to their rhetorical purpose and function rather than their lexico-grammatical structure. Explicitly teaching rhetorical

patterns across a variety of genres, through data-informed visualizations from DocuScope Classroom, may prime students to see relationships between writing tasks they encounter, enabling meaningful learning transfer. Within the classroom, we boost the potential for deep learning when we expose students to varieties of tasks within an integrated writing curriculum that emphasizes the range of skills and language patterns necessary for producing communication in the 21st century.

Acknowledgement

This work was supported by the Andrew W. Mellon Foundation (Award Number: 1712-05205).

Author Biographies

Danielle Wetzel is Teaching Professor of English at Carnegie Mellon University, where she directs the Writing & Communication Program. She works mainly on teacher training, curriculum design, and assessment, specifically for first-year writing. With the DocuScope team, she has piloted different versions of the DocuScope Classroom platform with her first-year writing students.

David West Brown is an Associate Teaching Professor of English at Carnegie Mellon University, where he teaches in the English and Statistics & Data Science Departments. He also directs research and assessment in the Writing & Communication Program.

Necia Werner is Associate Teaching Professor of English at Carnegie Mellon University, where she directs the undergraduate professional and technical writing program and the foundational professional and technical writing curriculum for nonmajors. Her research mainly focuses on the question of how we communicate expert knowledge to nonexpert audiences. She is the immediate past president of the IEEE Professional Communication Society.

Suguru Ishizaki is Professor of English at Carnegie Mellon University, where his current research focuses on computer-assisted rhetorical analysis and technology-enhanced writing instruction. Before his current appointment, he was a senior staff engineer at Qualcomm and was on the faculty at Carnegie Mellon's School of Design. He is the past president of the IEEE Professional Communication Society.

David Kaufer is Mellon Professor of English at Carnegie Mellon University. His research interests are rhetorical theory, composition theory, and computational approaches to both. With Suguru Ishizaki, he co-invented the DocuScope Text Analysis and Visualization environment and was responsible for the dictionary curation work on that project.

References

Alsop, S., & Gardner, S. (2019). Understanding attainment disparity: The case for a corpus-driven analysis of the language used in written feedback information to students of different backgrounds. *The Journal of Writing Analytics*, *3*, 38-68.

Andrade, H. L., Cizek, G. J., & Bennett, R. E. (2019). Preface. In H. L. Andrade, R. E. Bennett, & G. J. Cizek (Eds.), *Handbook of formative assessment in the disciplines* (pp. ix-xi). Taylor & Francis.

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice*, *48*(1), 12-19. https://www-jstor-org.proxy.library.cmu.edu/stable/pdf/40071571.pdf?refreqid=excelsior%3Af0bf1e2cc10ce9786c24a5b49ec23f25

Anson, C. M., & Moore, J. L. (Eds.). (2016). *Critical transitions: Writing and the question of transfer*. The WAC Clearinghouse; University Press of Colorado. https://doi.org/10.37514/PER-B.2016.0797

Aull. L. (2015). *First-year university writing: A corpus-based study with implications for pedagogy.* Palgrave Macmillan.

Aull, L. (2018). Corpus analytic tools: Constructing and understanding student writing assessment. *Assessing Writing*, *38*, 37-39.

Aull, L. (2019). Linguistic markers of stance in upper-level student writing. *Written Communication*, *36*(2), 267-295.

Bawarshi, A. S., & Reiff, M. J. (2010). *Genre: An introduction to history, theory, research, and pedagogy*. Parlor Press; The WAC Clearinghouse.

Bazerman, C., Applebee, A. N., Berninger, V. W., Brandt, D., Graham, S., Jeffrey, J. V., Matsuda, P. K., Murphy, S., Rowe, D. W., Schleppegrell, M., & Wilcox, K. C. (2018). *The lifespan development of writing*. National Council of Teachers of English.

Beaufort, A. (2007). *College writing and beyond: A new framework for university writing instruction*. Utah State University Press.

Biber, D. (1991). *Variation across speech and writing*. Cambridge University Press.

Biber, D. (1992). The multi-dimensional approach to linguistic analyses of genre variation: An overview of methodology and findings. *Computers and the Humanities*, *26*(5-6), 331-345.

Brown, D. (2020, February). *Using DocuScope for describing disciplinary and task variation* [Paper presentation]. The 9th Annual International Conference on Writing Analytics, St. Petersburg, FL, United States.

Brown, D. W., & Aull, L. L. (2017). Elaborated specificity versus emphatic generality: A corpus-based comparison of higher- and lower-scoring Advanced Placement exams in English. *Research in the Teaching of English*, *51*(4), 394-417.

Calkins, L., Ehrenworth, M., & Akhmedjanova, D. (2019). Creating formative assessment systems in the teaching of writing, and harnessing them as professional development. In H. L. Andrade, R. E. Bennett, & G. J. Cizek (Eds.), *Handbook of formative assessment in the disciplines* (pp. 173-206). Taylor & Francis.

Cizek, G. J., Andrade, H., & Bennett, R. E. (2019). Formative assessment: History, definition, and progress. In H. L. Andrade, R. E. Bennett, & G. J. Cizek (Eds.), *Handbook of formative assessment in the disciplines* (pp. 3-17). Taylor & Francis.

Conrad, S. (2017). The use of passives and impersonal style in Civil Engineering writing. *Journal of Business and Technical Communication*, *32*(1), 38-76.

Cotos, E., Huffman, S., & Link, S. (2015). Furthering and applying move/step constructs: Technology-driven marshalling of Swalesian genre theory for EAP pedagogy. *Journal of English for Academic Purposes*, *19*, 52-72.

Cotos, E., Huffman, S., & Link, S. (2017). A move/step model for methods sections: Demonstrating rigour and credibility. *English for Specific Purposes*, *46*, 90-106.

DeJeu, E., & Brown, D. (2021). A computational-rhetorical analysis of disciplinary variation in student writing. [Manuscript in preparation]. Department of English, Carnegie Mellon University.

Deng, Q., Hine, M. J., Ji, S., & Sur, S. (2018). Inside the black box of dictionary building for text analytics: A design science approach. *Journal of International Technology and Information Management*, *27*(3),119-159.

Geisler, C. (1994). *Academic literacy and the nature of expertise: Reading, writing, and knowing in academic philosophy*. Lawrence Erlbaum.

Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In C. Geertz (Ed.), *The interpretation of cultures: Selected essays by Clifford Geertz* (pp. 3-30) . Basic Books.

Gere, A. R., Knutson, A. V., & McCarthy, R. (2018). Rewriting disciplines: STEM students' longitudinal approaches to writing in (and across) the disciplines. *Across the Disciplines*, *15*(3), 63-75. https://wac.colostate.edu/docs/atd/trans/gereetal2018.pdf

Gray, B. (2013). More than discipline: uncovering multi-dimensional patterns of variation in academic research articles. *Corpora*, 8, 153-181.

Hardy, J. A., & Römer, U. (2013). Revealing disciplinary variation in student writing: A multi-dimensional analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, *8*(2), 183-207.

Hayes, H., Ferris, D. R., & Whithaus, C. (2016). Dynamic transfer in first-year writing and "writing in the disciplines" settings. In C. Anson & J. Moore (Eds.), *Critical transitions: Writing and the question of transfer* (pp. 181-213). The WAC Clearinghouse; University of Colorado Press.

Helberg, A., Poznahovska, A., Ishizaki, S., Kaufer, D., Werner, N., & Wetzel, D. (2018). Teaching textual awareness with DocuScope: Using corpus-driven tools and reflection to support students' written decision-making. *Assessing Writing*, *38*, 40-45.

Hillocks, G. (2002). *The testing trap: How state writing assessments control learning*. Teachers College Press.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177). Association for Computing Machinery.

Ishizaki, S., & Kaufer, D. (2011). DocuScope: Computer-aided rhetorical analysis. In P. McCarthy & C. Boonthum (Eds.), *Applied natural language processing and content analysis: Advances in identification, investigation, and resolution* (pp. 276-297). IGI Global.

Jönsson, A., & Eriksson, U. (2019). Formative assessment in higher education: An example from astronomy. In H. L. Andrade, R. E. Bennett, & G. J. Cizek (Eds.), *Handbook of formative assessment in the disciplines* (pp. 146-169). Routledge.

Kaufer, D., & Butler, B. S. (2000). *Designing interactive worlds with words: Principles of writing as representational composition* (1st ed.). Routledge.

Kaufer, D., Ishizaki, S., Butler, B. S., & Collins, J. (2004). *The power of words: Unveiling the speaker and writer's hidden craft*. Routledge.

Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research*, *1*(1), 1-26.

Moss, P. A., Pullin, D., Haertel, E. H., Gee, J. P., & Young, L. J. (Eds.). (2008). *Assessment, equity, and opportunity to learn.* Cambridge University Press.

National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st centur*y. The National Academies Press. https://doi.org/10.17226/13398

Oliveri, M. E., Mislevy, R. J., & Slomp, D. H. (2021). Principled development of Workplace English Communication Part 1: A sociocognitive framework. *The Journal of Writing Analytic*s, *5*.

Oliveri, M. E., Slomp, D. H., Elliot. N., Rupp, A. A., Mislevy, R. J., Vezzu, M., Tackitt, A., Nastal, J., Phelps, J., & Osborn, M. (2021). Introduction: Meeting the challenges of Workplace English Communication in the 21st century. *The Journal of Writing Analytics*, *5*.

Oliveri, M. E., Slomp, D. H., Rupp, A. A., & Mislevy, R. J. (2021). Principled development of Workplace English Communication Part 2: Expanded Evidence-Centered Design and Theory of Action frameworks. *The Journal of Writing Analytics*, 5.

Patchan, M., Charney, D., & Schunn, C. (2009). A validation study of students' end comments: Comparing comments by peers, a writing instructor, and a content instructor. *Journal of Writing Research*, *1*(2), 124-152.

Reese, A. N., Rachamalla, R. R., Rudniy, A., Aull, L., & Eubanks, D. (2018). Contemporary peer review: Construct modeling, measurement foundations, and the future of digital learning. *The Journal of Writing Analytics*, *2*, 96-137.

Ringler, H., Klebanov, B. B., & Kaufer, D. (2018). Placing writing tasks in local and global contexts: The case of argumentative writing. *The Journal of Writing Analytics*, *2*, 34-77.

Ross, V., & LeGrand, R. (2017). Assessing writing constructs: Toward an expanded view of inter-reader reliability. *The Journal of Writing Analytics*, *1*, 227-275.

Sardinha, T. B., & Pinto, M. V. (Eds.). (2014). *Multi-dimensional analysis, 25 years on: A tribute to Douglas Biber* (Vol. 60). John Benjamins Publishing Company.

Slomp, D., Burke, T., Neamtu, K., Hagen, L., Van Ham, J., Miller, K., & Dupuis, S. (2018). Scaffolding for independence: Writing-as-problem-solving pedagogy. *English Journal*, *108*(2), 84-94.

Solacci, L. B., & Pereira, M. G. (2004). The introduction, methods, results, and discussion (IMRAD) structure: A fifty-year survey. *Journal of Medical Libraries Association*, *92*(3), 364-367.

Swales, J. M. (1981/2011). *Aspects of article introductions*. University of Michigan Press.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings.* Cambridge University Press.

Trachsel, M. (1992). *Institutionalizing literacy: The historical role of college entrance examinations in English.* Southern Illinois University Press.

WAC Clearinghouse. (n.d.). *What is writing in the disciplines?* https://wac.colostate.edu/resources/wac/intro/wid/

Wardle, E. (2009). "Mutt genres" and the goal of FYC: Can we help students write the genres of the university? *College Composition and Communication*, *60*(4), 765-789.

Young, L., & Soroka, S. (2012). Affective news: The automated coding of sentiment in political texts. *Political Communication*, *29*(2), 205-231.