Understanding Graduate Writers' Interaction with and Impact of the Research Writing Tutor during Revision

Elena Cotos, Sarah Huffman & Stephanie Link

Iowa State University, Iowa | USA
Oklahoma State University, Oklahoma | USA

Abstract: Teaching the craft of written science communication is an arduous task that requires familiarity with disciplinary writing conventions. With the burgeoning of technological advancements, practitioners preparing novice research writers can begin to augment teaching and learning with activities in digital writing environments attuned to the conventions of scientific writing in the disciplines. The Research Writing Tutor (RWT) is one such technology. Grounded in an integrative theoretical framework, it was designed to help students acquire knowledge about the research article genre and develop research writing competence. One of its modules was designed to facilitate revision by providing different forms of automated feedback and scaffolding that are genre-based and discipline-specific. This study explores whether and how the features of the RWT may impact revision while using this module of the tool. Drawing from cognitive writing modeling, this study investigates the behaviors of a multidisciplinary group of 11 graduate-student writers by exploring how they interacted with the RWT's features and how this interaction may create conditions for enhanced revision processes and text modifications. Findings demonstrate promising potential for the use of this automated feedback tool in fostering writers' metacognitive processing during revision. This research adds to theory on cognitive writing models by acknowledging the evolving role of digital environments in writing practices and offering insights into future development of automated tools for genre-based writing instruction.

Keywords: genre, research writing, revision, automated rhetorical feedback

journal of WRITING RESEARCH

Cotos, E., Huffman, S., & Link, S. (2020). Understanding graduate writers' interaction with and impact of the Research Writing Tutor during revision. *Journal of Writing Research, 12(1),* 187-232. https://doi.org/10.17239/jowr-2020.12.01.07

Contact: Elena Cotos, Iowa Stae University, 1137 Pearson Hall, 505 Morrill Rd., Ames, IA 50011-2103 | USA – ecotos@iastate.edu

1. Introduction

An important but challenging goal of writing instruction in graduate education contexts is to help students develop research writing competence and to aid their transition from peripheral to full participation in the scholarly discourse of their discipline. Research writing reflects the persuasive nature of knowledge creation and requires crafting a scientific argument in rhetorically and linguistically explicit ways (Swales, 1990) that are appropriate to the expectations of the discourse community, a social structure with an established system of practices (Giddens, 1979). Rhetoric is important, as research writers intend to persuade readers of the need for and noteworthiness of their scientific endeavor in ways that are established and conventionalized within disciplines. Therefore, teaching and assessing research writing requires familiarity with the disciplinary conventions of research genres, which makes it an onerous task for writing instructors. Teachers, as well as their students, would thus greatly benefit if assisted by genre-based and discipline-specific writing tools.

In recent decades, the number and variety of new generation writing tools for university as well as school-aged students have proliferated across teaching and learning contexts. Representative examples such as the AcaWriter (Knight et al., 2020), C-SAW (Benetos & Bétrancourt, 2020), Inputlog (Vandermeulen, Leijten, & Van Waes, 2020), and MI Write (Palermo & Wilson, 2020) are described in this special issue. Allen, Jacovina, and McNamara (2016) view digital writing tools as classifiable into systems that provide automated scoring, automated evaluation/feedback, and intelligent tutoring. Adopting and adapting this classification, Strobl et al. (2019) identified 89 digital academic writing tools developed to support writing in different educational settings and conducted a comprehensive review of select 44 that supported student writing in higher education. Of these, only a handful were designed for research-related writing genres. The few existing ones originated in different fields (e.g., applied linguistics and writing analytics) and offer their own unique features. For example, the Academic Word Suggestion Machine (Mizumoto, Sawako, & Imaob, 2017) automatically suggests word combinations frequently occurring in certain sections of the research article (e.g., "the article concludes with") as the student types in a word. The Thesis Writer (Rapp & Kauf, 2018) takes a process-based approach, facilitating students' writing process from the beginning to the completion of their thesis, also offering features that afford collaboration. Another tool, not included in Strobl et al.'s (2019) review but featured in this issue, AcaWriter (Abel, Kitto, Knight, & Buckingham Shum, 2018; Knight et al., 2020) includes a rule-based parser (Sándor, 2007) that enables automatic analysis of students' texts, including research article Introduction sections, and provides feedback on rhetorically salient sentences (e.g., a sentence may contrast ideas). A

shared characteristic of all these tools is that some of their scaffolding features rely on disciplinary corpora representative of research writing.

This paper presents one more exemplar of corpus-based tools, an automated evaluation/feedback system for the research article genre – the Research Writing Tutor (RWT) (Cotos, 2016). Informed by research on published scientific discourse in a wide range of disciplines (Cotos, Huffman, & Link, 2015), the RWT offers various affordances for learning, analyzing, and producing the Introduction, Methods, Results, and Discussion/Conclusion (IMRD/C) components of the research article (RA). Its three modules include: 1) instructional materials comprehensively describing the RA genre conventions, 2) pedagogically-mediated corpora of RAs in thirty disciplines, and 3) discipline-specific automated feedback on student writing with on-demand scaffolding. All three modules are described in Section 3; the latter module, called Analyze My Writing, was the one used in this study.

The RWT is an online platform developed for graduate student writers and their teachers, aiming to complement formal academic writing instruction. Prior to its current implementation, the RWT was first prototyped and empirically evaluated in the context of graduate writing courses at Iowa State University. The prototype, called Intelligent Academic Discourse Evaluator (IADE), was designed only for the Introduction section and tested with English as a second-language writers. Its design was guided by a conceptual framework that aligned teaching and learning needs with tenets rooted in learning and language theories (Cotos, 2009). This allowed for designing features in view of needs-based and theoretically-informed assumptions about how to best create conditions necessary for the development of research writing competence. Research investigating students' use of IADE provided support to the initial assumptions, showing that the tool's features could facilitate genre awareness and learning, and could contribute to improvement in Introduction writing (Cotos, 2011, 2012). Importantly, the studies also revealed the potential of this pilot genre-based technology to enhance the cognitive and socio-disciplinary dimensions of RA writing (Cotos, 2014). Consequently, the theoretical grounding for the design and use of the RWT builds on the conceptual framework of its prototype and also comprises additional theoretical underpinnings.

In the next section, we explain how different theories merged into a broader overarching framework and link the theoretical tenets to the specific features they informed, following with a description of what the RWT's features look like and what they are intended for in Section 3. We then introduce the study in Section 4, briefly explaining the theoretical perspective that guided this research and describing the methods we employed. Having then reported and discussed the results in Sections 5 and 6, we conclude with indications for future developments.

2. Theoretical framework

The theoretical grounding guiding the design of the RWT concentrates on the desired learning processes and outcomes vis-à-vis the needs of teachers and students. In short, the teaching needs included enacting an explicit pedagogy (Hyland, 2007) for teaching the RA genre, which would enable students to deconstruct texts and tease out "for own critical appreciation and understanding, how a writer as maker or fabricator has gone about constructing and shaping that text" (Badley, 2009, p. 213). Explicit genre-based instruction can give teachers the means needed to more effectively acquaint students with conventional textual instantiations of the RA genre and to cultivate an awareness of the nature of scientific argumentation as established in the disciplines. This is essential to fostering students' formal, rhetorical, and procedural aspects of genre knowledge (Tardy, 2009) and to developing their research writing competence. The students' ultimate goal was to develop an ability to create genre artifacts that were congruent with the values of the target socio-disciplinary practice. For that, they needed exposure to authentic disciplinary discourse, directions for how to discern the writing norms of their discourse community, guided writing practice, individualized feedback, and productive interaction. Creating opportunities to cater to all these needs required drawing on an integrated theoretical framework that would inform the design of technological features with affordances for both socio-disciplinary and cognitive dimensions.

2.1 Socio-disciplinary dimension

The socio-disciplinary dimension is enfolded by genres. Genres are generally viewed as responses to social interactions, which are realized as recurring textual representations that are ingrained with conventions established by parent discourse communities (Polio & Williams, 2011). Considering the need for explicit pedagogy, which to a great extent originated from linguistic perspectives (Hyland, 2003), the main underpinnings of the RWT draw on English for Specific Purposes and Systemic Functional Linguistics. These are informative both as theoretical perspectives and pedagogical approaches.

In English for Specific Purposes, genres are theorized as texts types that are organized in relation to communicative purposes. Genres are characterized by recognizable rhetorical structures, which strategically package content in ways that are purposefully and routinely employed by the members of a discourse community (Swales, 1990). These structures, known as *moves*, are segments of a text that attain particular communicative goals characteristic of a particular genre. Moves are composed of *step*s or strategies that writers use to accomplish the communicative goals. Writers make specific language choices that serve as explicit signals indicating the functions of the strategies used. In simple terms:

Texts of a genre have parts (IMRD/C sections in the research article)
→ parts have communicative goals (moves)
→ goals are achieved by strategies (steps)
→ strategies use typical language (functional linguistic signals).

Appendix A contains definitions and examples of the moves and steps of the Introduction section of a research article, which originated from Swales' (1981) Create-A-Research-Space model. The communicative goals of one of the moves in this model, 'Establishing a niche,' is to expose issues in the current state of the art on a topic in order to demonstrate the need for the new study. This move can be realized by using different steps: 'indicating a gap,' 'highlighting a problem,' 'raising general questions,' 'proposing general hypotheses,' and 'presenting a justification.' Some examples of language choices that function to highlight a problem, for instance, include negative quantifiers (no, little, none of, very few, neither… nor), verbs (fail, lack, overlook), adverbs (rarely, scarcely, barely, hardly), nouns (failure, limitation, gap, dearth, lack), or adjectives (inconclusive, misleading, scarce, elusive, limited, questionable).

Similarly, Systemic Functional Linguistics views genres as "staged goal-oriented social processes" (Martin, 1993, p. 13), also affirming that textual patterns underlie socially recognized communicative functions. The theoretical focus here is on how the linguistic system serves as social semiotics to create meaningful communication. Meaning is inseparable from form. As Halliday and Hasan (1989) put it, meanings are encapsulated in texts "through a systematic relationship between the social environment on the one hand, and the functional organization of language on the other" (Halliday & Hasan, 1989, p.11). In genres, it is the verbal strategies, or linguistic choices, that function to accomplish social purposes (Martin, 1985, p. 251).

The RWT leverages the complementarity of these two perspectives to operationalize the systematic relation between discourse structure, rhetorical intent, and functional language. Hyland (2003) recommends placing "an explicit focus on the ways texts are organized and the language choices that users must make to achieve their purposes in particular contexts" (p. 75) as the starting point of explicit pedagogy for developing genre knowledge (particularly formal and rhetorical, as defined by Tardy (2009)), which is why the constructs of move/step and functional language govern most of RWT's features. The tool allows for deconstructive analysis, evaluation, and interpretation of texts geared towards the discovery of discipline-specific structural, rhetorical, and linguistic characteristics. With that, the RWT is aimed to help students grasp the connection between texts and the socio-disciplinary practice they realize, and to help them learn how to appropriately use linguistic features as functional choices to express particular meanings in larger units of genre discourse.

2.2 Cognitive dimension

Accounting for the cognitive dimension is important for any learning technologies because the acquisition of any skill is, in essence, a form of learning that triggers cognitive activity. Research writing as a skill entails engaging in highly reflective processes to produce texts that effectively present the outcomes of empirical inquiry in a re-constructive synthesis of existing and newly acquired scientific knowledge. To develop this complex ability, novice writers need extensive practice and feedback. The role of these concepts is emphasized by cognitive writing, skill acquisition, and language acquisition theories.

Cognitive models of writing theorized over the decades (Butterfield, Hacker, & Albertson, 1996; Chanquoy, 2009; Chenoweth & Hayes, 2001; Fitzgerald, 1987; Galbraith, 2009; Flower & Hayes, 1981; Hayes, 2004; Hayes, 2012; Hayes, Flower, Schriver, Stratman, & Carey, 1987; Horning, 2002; McCutchen, Teske, & Bankston, 2008) are important in considering the complex cognitive mechanisms activated in the process of research writing. Overall, writing is viewed as a goal-oriented process with recursive cognitive activities such as planning, translating, and revising. Revising is a particularly intense cognitive operation, for it involves comparing the intended and instantiated text to identify discrepancies, diagnosing what and how in the text should be changed, and operating by selecting appropriate strategies to make the desired text modifications (Bereiter & Scardamalia, 1983; Fitzgerald, 1987). It is these revision processes that distinguish expert and novice writers, as the latter are less able to detect, diagnose, and rectify the mismatch between the intended mental and actual representation of their text.

The Skill Acquisition Theory also centers on cognitive representation, but within a more general scope. This theory postulates that the acquisition of different skills bares similarity due to a cognitive architecture, a production system model that underlies the cognitive processes activated when learning to perform a task.[1] This cognitive architecture consists of three consecutive stages of development – presentation, practice, and production (Byrne, 1986).[2] At the presentation stage, learners acquire declarative knowledge by attending to explicit demonstration or explanation of aspects of the task. The practice stage is to proceduralize declarative knowledge into behavior, so that 'knowledge that' turns into 'knowledge how' (DeKeyser, 2007). Repeated and deliberate practice, during which the learner receives or monitors feedback, leads to determining and selecting the most efficient procedures for performing the task. Extensive practice ultimately enables gradual automatization of procedures for task production and improved task performance, with a greatly reduced error rate and need for external support.

For non-native speakers of a language, writing is not just a skill – it is a language skill. Therefore, theoretical insights from second language acquisition must enter the scene if writing tools are to also be used by second language writers. In particular, the Interaction Hypothesis in Second Language Acquisition Theory links

interaction and learning, highlighting cognitive concepts such as noticing, working memory, and attention (Gass & Mackey, 2007). These internal processes play a significant role as the learners encounter linguistic input, engage in interaction, and produce output. Feedback is a crucial aspect of interaction because it often supplies so-called negative evidence, which helps the learners notice a problem in their language production, i.e. a mismatch between the input and their language use (Mackey, 2006; Schmidt & Frota, 1986). Noticing of negative evidence prompts learners to modify their linguistic form to produce more comprehensible output (Swain, 1985) and convey the intended meaning more effectively.

These three theoretical perspectives are grounded in different fields and schools of thought. Nevertheless, they share some similar ideological interpretations while each providing special guidance in terms of how to create springboard conditions for helping students develop genre knowledge and research writing competence. The RWT leverages this theoretical basis in an integrated principled design of various features. Most heavily it focuses on individualized feedback, which should assist students during the practice stage of skill acquisition, help them detect and diagnose shortcomings in the written text compared to their intended mental representation, and trigger their noticing of negative evidence in language use when conveying specific functional meanings. Iterative revision with scaffolding for metarhetorical, metastrategic, and metalinguistic awareness is the second major focus, as revision practice is what should help the novice writers learn how to develop most efficient strategies and select most appropriate language choices, and move from the proceduralization to the automatization of writing skills to produce expert-like RA genre discourse.

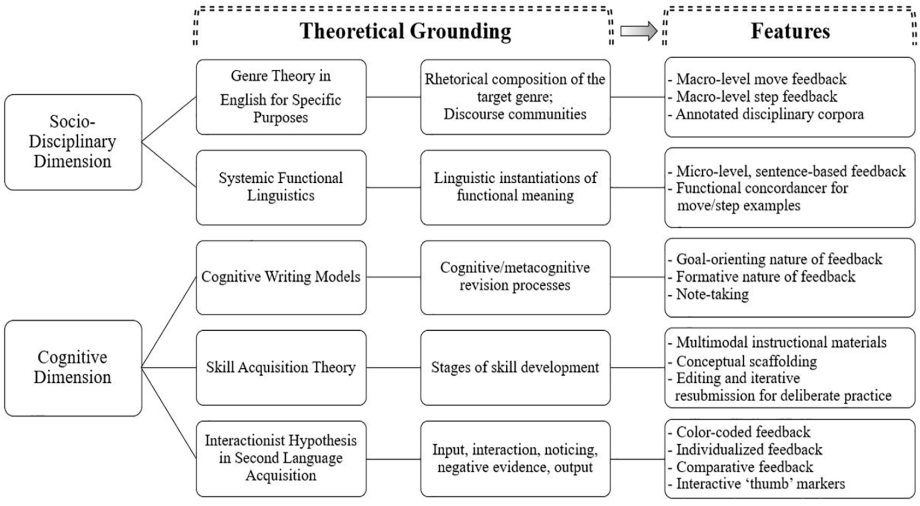

Figure 1. Theoretical tenets underpinning RWT's features.

Figure 1 shows the precise connections between theoretical concepts operationalized in the RWT's features.

3. The Research Writing Tutor (RWT)

The RWT is a web-based platform that offers a number of features with affordances for budding researchers to help them recognize and apply normative conventions of the RA genre in their disciplines. Similar to other writing platforms (e.g., AcaWriter, Criterion, MI Write (successor of Project Essay Grade Writing), MyAccess, Writing Mentor, Cambridge English Write & Improve), the RWT provides instant automated feedback. Like C-SAW, it renders multiple visual representations. What makes this tool unique is the rhetorical nature of the feedback given through its Analyze My Writing module and tailored to each of the IMRD/C sections in a range of disciplines. As noted in Figure 1, the RWT generates different types of individualized feedback. It affords students the ability to concentrate on the communicative effectiveness of the discourse in their IMRD/C section drafts and to adopt writing norms common to the RA genre in their discipline through the provision of both macro-level and micro-level feedback, both of which apply Swales' (1981) rhetorical constructs of moves and steps.

When the students submit a written draft of a section of the RA, the RWT provides immediate macro-level feedback based on the occurrence of moves and steps, at the same time visualizing the move composition of the draft. Figure 2 exemplifies the macro-level feedback on Introduction moves (i.e., 'Establishing a Territory,' 'Identifying a Niche,' and 'Addressing the Niche'). This type of feedback takes different forms, one of which is the color-coded representation of the moves for the purpose of input enhancement. Another form of this type of feedback is graphical/numerical, presented as range bars (next to the color-coded text) and pie charts (below the color-coded text) that are based on a comparison of the draft with the rhetorical functions identified in a corpus of published disciplinary articles. This comparison adds a goal-orientation quality to the macro-level move feedback by helping the students see if there is a potential mismatch between move and step distributions in their draft and in the larger context of articles in their fields.

The range bars are expandable to provide macro-level feedback on steps, being more specific to the rhetorical functions of the given move. The expanded portion indicates the use of steps in the draft. Figure 2 above exemplifies the feedback for the three steps of move 1: 'providing general background,' 'reviewing previous research,' and 'claiming centrality.' Viewing this analyzed draft through the lens of macro-level move and step feedback should help the students conceive of the relation between the functional components of their drafts and the overall communicative goals of the section.

The macro-level feedback on steps is accompanied by on-demand conceptual scaffolding, as the students can click on or hover over the question mark icon next

Figure 2. Macro-level feedback on moves and steps

to each move and access respective descriptive definitions. By clicking on Learn More, the students are directed to another module of the RWT, called Understand Writing Goals, which offers multi-modal instructional materials including short video-lectures and detailed descriptions with examples of each move and step of each of the IMRD/C sections. Brief descriptions of the steps can be accessed by hovering over Learn More. In Figure 2, an arrow points to the pop-up definition of the 'claiming centrality' step.

Another scaffolding feature is provided through the See Examples link, which brings up a page from the concordancer in the Explore Published Writing module of the RWT. This corpus-based module is used in data-driven learning class activities (see Cotos, Link, & Huffman, 2017a), as it contains RA corpora that are annotated for moves and steps in 30 disciplines and focuses students' awareness on the macro-level rhetorical composition of published RAs in respective disciplines. Figure 3 displays a screenshot of examples of the 'claiming centrality' step of move 1 in Horticulture Introductions. The concordance lines maintain the same color-code for input enhancement as in the feedback. As it is shown in the lower part of Figure 3, the examples can be seen in the textual context, both orienting the students to

where they are situated in original texts and enabling them to discover patterns in the annotated disciplinary corpora. Interaction with the contextualized examples of authentic, published discourse composed by more experienced research writers is expected to draw students' attention to the linguistic instantiations of the steps in building up argumentation, and the layered functionality of texts within their social environment (Halliday & Hasan, 1989). Such exposure to the RA genre and hands-on engagement with field-relevant writing can serve as a means of ushering students' socialization into their disciplinary discourse communities.

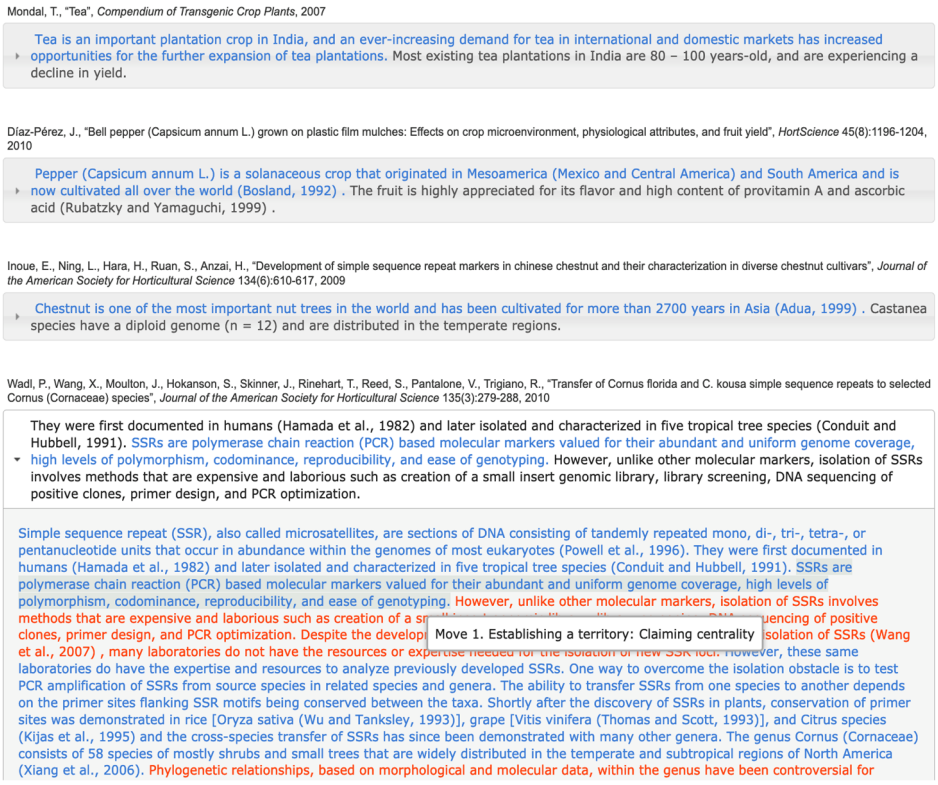

Figure 3. Functional concordancer feature.

Additionally, the RWT offers micro-level feedback on steps in the form of sentence-based prompts. When clicking on or hovering over a color-coded sentence, the students are shown a prompt suggesting the function of the step that the student may be attempting with that sentence. In Figure 4, the feedback suggests that the sentence clicked on may be 'highlighting a problem' (a step of move 2).

Figure 4 also depicts two other integral features within the Analyze My Writing module. One feature allows the students to act upon the sentence-based prompts

by interacting with 'thumb' markers, which motivates them to think whether indeed a prompt reflects what they intended to convey or not, and then to take notes for further revision.

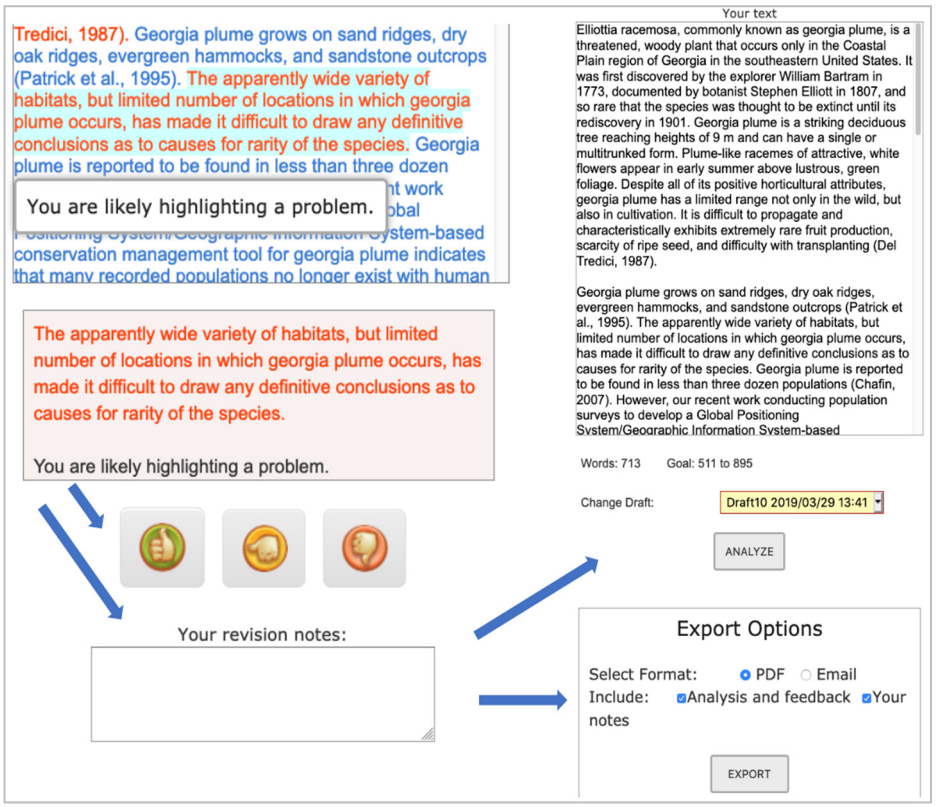

Figure 4. Note-taking for revision, editing, and iterative resubmission

The note-taking feature allows them to export their notes through different combinations of options (e.g., the analysis and feedback plus revision notes as a PDF document sent via email). Another feature allows for iterative resubmission. After making modifications in the editing box, the students can click on the Analyze button and re-submit the text they modified in response to one or more types of automated feedback. This feature enables them to automatically analyze their draft and get updated feedback as many times as needed.

4. Empirical study

The effectiveness of the RWT has been previously investigated from different perspectives, including usability and usefulness (Ramaswamy, 2012), the fit of its features for students with targeted learner characteristics (Cotos & Huffman, 2013), performance of the feedback engine (Cotos & Pendar, 2016), and students' use of corpus-based features in the Explore Published Writing module for developing RA genre knowledge through data-driven learning tasks (Cotos et al., 2017a). This study continues the latter strand of research by focusing on students' use of the features in the Analyze My Writing module as they worked on their Introduction drafts. Specifically, the purpose was to investigate their impact on revision.

4.1 Theoretical orientation

Given that propositions from the cognitive writing models were included in the theoretical framework guiding the design of the RWT, in this study we set on exploring the role of technology in fostering metacognitive activity during revision, as it needs to be theoretically and practically transparent. Cognitive writing models have evolved; yet, computer-mediated writing support has not been explicitly accounted for in models theorizing revision, especially considering writing for specific purposes to specific target audiences – although important advancements have been made by researchers working on automated essay scoring (e.g., Deane (2013) and Deane, Quinlan, Odendahl, Welsh, & Bivens-Tatum (2008)) and intelligent tutoring for writing strategies (e.g., Roscoe, Snow, & McNamara, 2013)).

Drawing from Butterfield, Hacker, and Albertson (1996), we put forth an augmented version of the revision model given in Figure 5. We position the RWT's feedback and just-in-time conceptual scaffolding (Jackson, Krajcik, & Soloway, 1998; Schwarz & Bransford, 1998) about rhetorical and disciplinary conventions, as well as iterative resubmission, within the environment (shaded box in Figure 5).

In their model, Butterfield et al. (1996) consolidate the essential environmental and cognitive-metacognitive aspects of revision. The environment part captures rhetorical problems, including problems related to the topic, audience, and importance of the text to be revised, as well as the actual text. The actual text may range from distinct lexical units to extended discourse, such as a full draft. The cognitive/metacognitive dimension is rendered through processes in the long-term memory and working memory. The former represents revisers' automatic monitoring and control of their knowledge and understanding of strategies for revision, while the latter is where deliberate revision processes such as problem representation, evaluation, strategy selection, and execution take place. Since Butterfield et al. proposed their model of revision, advancements in technology have changed revisers' experience. Genre-based technology, like the RWT, provides an environment for creating a new representation of how an actual text maps onto genre standards and poses different rhetorical problems for deliberation

in the revisers' working memory. Consequently, reflecting this environmental difference in cognitive modeling of revision and investigating it is imperative.

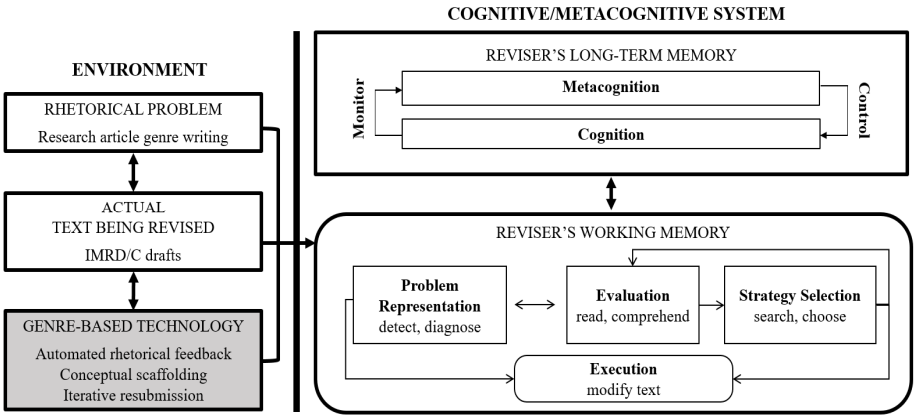

Figure 5. Augmented text revision model based on Butterfield, Hacker, and Albertson (1996).

4.2 Purpose and research questions

In this study, we explored the link between the technology-altered environment and the revision processes occurring in the working memory (Fig. 5) by examining whether and how the features of the RWT, as an external environmental factor, may impact revision. The research questions (RQs) we sought to answer are as follows:

- RQ1. How do novice research writers interact with the RWT's feedback and scaffolding features during revision?
- RQ2. In what ways may their interaction with the RWT's feedback and scaffolding features impact their revision processes?
- RQ3. In what ways may their interaction with the RWT's feedback and scaffolding features impact their execution of text modifications?

4.3 Participants and instructional context

Data were obtained from 11 participants who were all pursuing a master's or doctoral degree in different disciplines (see Table 1). The participants were enrolled in an advanced academic writing course offered at Iowa State University in the United States. The course focused on preparing a publishable manuscript, with a prerequisite requirement that the students had to have completed a research study. The teaching practices were anchored in genre analysis research and genre-based pedagogy in English for Specific Purposes, using corpus-based materials and activities both within and outside of the RWT platform (see Cotos et al., 2017a). In

this study, we focus on the students' use of the tool's Analyze My Writing module to revise their RA Introduction drafts during the eighty minutes of a class period.

Table 1. Participant demographics (N = 11)

Participant	Gender	Age	Language Background*	Discipline
P1	F	26-30	NNS	Applied Linguistics
P2	F	41+	NNS	Statistics
P3	M	41+	NS	Materials Science & Engineering
P4	F	26-30	NNS	Applied Linguistics
P5	F	41+	NNS	Education
P6	F	26-30	NS	Chemistry
P7	F	36-40	NS	Food Science
P8	M	21-25	NS	Mechanical Engineering
P9	F	26-30	NNS	Materials Science & Engineering
P10	M	31-35	NNS	Chemical Engineering
P11	M	26-30	NNS	Sociology

*NS = Native English speaker, NNS = Nonnative English Speaker

Only one section of the course was offered at the time of this study, which means that it was not possible to recruit students from another section to participate as a control group. It was also not possible to divide the 11 participants we recruited into experimental and control groups because of the low number of students enrolled and also for pedagogical and fairness reasons (highlighted by the Institutional Review Board). Therefore, our study design included both qualitative and quantitative analyses of different types of data obtained from this small sample of the target population.

The class met twice a week, with each session lasting for an hour and twenty minutes. Data were collected during one class period, most of which was devoted to students' using the RWT to revise their first draft (assigned two days before this class). At the beginning, there was a demo of the Analyze My Writing module. This did not require much learner training because the students were familiar with other modules of the tool, which had features designed following similar formats.

4.4 Study design and data analysis

Our study employed a mixed-methods design (Cresswell & Clark, 2011), combining quantitative and qualitative analyses of process data (student-system interaction logs, on-task screen recordings), introspective data (stimulated recalls), and writing

product data (first and last drafts). We adopted a concurrent transformative strategy, which is characterized by the use of a theoretical perspective (revision model in Cognitive Writing Theory) reflected in our research questions and entails the collection of qualitative and quantitative data at the same time or in parallel. The integration of the data relates to the research questions and occurred mostly at the analysis phase (see Figure 6). The findings from both data collection methods were also integrated during interpretation, allowing to obtain an in-depth understanding of students' interactive behavior when revising with the RWT.

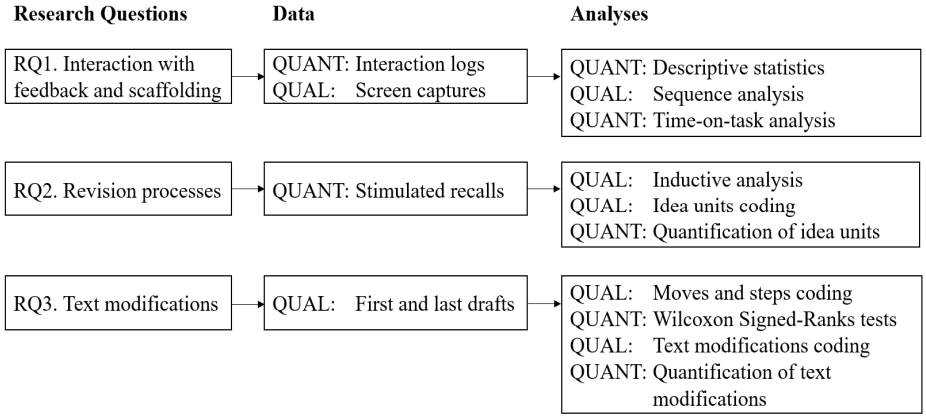

Figure 6. Study design.

Process data (student-RWT interaction logs and on-task screen capture) were included in the research design to observe how the students interacted with multi-level rhetorical feedback and scaffolding during revision (RQ1). They utilized the RWT for self-analysis of their Introduction draft by using the features of the Analyze My Writing module, making revisions based on the feedback, and resubmitting their revised drafts. The frequency of access to the features and the frequency of drafts submitted were obtained from the tool's database. These data were used to complement the screen captures, which were recorded using Apple Inc's QuickTime. Frequency tallies also gave a picture of general trends and patterns in user behavior (see Heift, 2002). Per Nix and Wylie's (2011) recommendations, descriptive statistics of the tallies were generated to render an overall understanding of participants' interaction behaviors. Qualitative analysis of students' screen captures adopted the sequence model, a task analysis-oriented, human-computer interaction observational technique (Phipps, Meakin, & Beatty, 2011). This allowed for analyzing individual students' data in terms of the sequence of their interactivity (as indicated by mouse movements) with particular features of the tool. Additional time-on-task analysis (Phipps, Meakin, & Beatty, 2011) was done to determine the length of time the students spent interacting with those features.

The screen capture recordings also served to collect introspective data (stimulated recalls), which allow researchers to access participants' thoughts as they engage in a task (Gass & Mackey, 2007). This type of data helped gain insight into the ways in which students' interactions with the RWT may impact the problem representation, evaluation, strategy selection, and execution revision processes (RQ2). Prior to each stimulated recall session, the screen captures were closely examined, and noteworthy behaviors were documented for more efficient and participant-tailored data elicitation. Each session was audio recorded for later transcription; then, transcripts were analyzed inductively by first coding according to prominent themes (as in Pujola, 2001; Nix & Wylie, 2010; Cotos, 2011). This coding process involved a preliminary reading where dominant recurring themes were identified. A second coding round aimed to substantiate the initial themes and to integrate any other phenomena occurring in participant responses. The last round of coding applied the refined categories to the entire stimulated recall dataset, the unit of coding being the idea unit (i.e., "a chunk of information which is viewed by the speaker/writer cohesively as it is given a surface form ... related ... to psychological reality for the encoder" (Kroll, 1977, p. 85)). The refined categories constituted a total of 6 themes (reflection, text processing, problem solving, strategizing, text production, and environment) and 21 sub-themes (frequencies further reported in Section 5.2).

Finally, writing product data in the form of first and last drafts of Introduction sections were collected to examine how the students executed text modifications appropriate for the genre (RQ3). The dataset contained 22 drafts from our 11 participants, which were processed using the RWT's automated analysis and feedback engine such that each sentence of each draft was labeled with one of the moves and one of the steps described in Appendix A.[3] Move and step codes for each sentence were counted per student and per draft to record the total number of move/step occurrences from first to last drafts, as this would help make interpretations about the rhetorical composition of their texts.

To determine whether there was a change in the rhetorical composition from first to last draft, step counts were used in a series of Wilcoxon Signed-Ranks tests, which are appropriate for within-groups comparisons of two dependent variables when the N-size is relatively small. Effect sizes were calculated by dividing the z-value by the square root of N, where N was the number of occurrences over the two drafts (Pallant, 2007). The effect sizes ranged from .12 to .63 and were assessed against criteria put forth by Plonsky and Oswald (2014), where .25 is a small effect, .40 is a medium effect, and .60 is a large effect. To further understand the text modifications effected at other levels, the first draft and last drafts were analyzed sentence by sentence, and the types of changes identified were quantified for comparison.

5. Results

5.1 Interaction with feedback and scaffolding features

The ways in which the students interacted with the RWT during revision were analyzed both quantitatively and qualitatively. Table 2 reports the descriptive statistics based on the frequency with which the students accessed particular features, as logged in the tool's database. It appears that our group of students preferred to primarily use the macro-level move feedback and the micro-level sentence-based feedback. Interactions with other features, such as note-taking and the concordancer, also occurred but were less common.

Note-taking and 'thumbs' are features linked in the system; in other words, to be able to take notes the student needs to click on one of the thumbs first. That is why these features were analyzed together. The 'thumbs up' option indicated their agreement with the feedback prompt on a given sentence (M = 1.36, Mdn = 0, SD = 3.64), 'thumbs down' indicated their disagreement (M = 6.36, Mdn = 5, SD = 5.22), and 'thumbs neutral' indicated that they partially agreed (M = 1.82, Mdn = 2, SD = 1.25). Students mostly took notes when they clicked on the 'thumbs down' and the 'neutral thumb' options, likely because they did not need to write notes about sentences they did not think had to be improved.

Table 2. Frequency of access logs in the RWT's database

Features	M	Md	SD
Macro-level move feedback	318.90	310	203.40
Macro-level step feedback	65.09	55	38.34
Micro-level sentence-based step feedback	116.72	119	63.63
Note-taking	9.55	9	6.06
Concordancer	4.54	3	3.50
Iterative resubmission (number of drafts)	3.54	2	4.98

With regards to iterative resubmission, there was variation in the number of drafts submitted for automated analysis. Figure 7 shows a breakdown of students according to drafts submitted and total mouse activity. Each point on the graph signifies an individual student participant. The X axis indicates the total number of drafts submitted by a given student; the Y axis shows how frequently the student accessed the features of the RWT. Noticeable in this figure are two outliers: Participant 2 submitting 5 drafts and clicking or hovering over the features at a high frequency (1,208), and Participant 4 submitting numerous drafts (18) and interacting at a frequency comparable with most other students (602). Despite these outliers, there appears to be a trend in this data showing that most of the students interacted with the RWT similarly, submitting up to 4 drafts within the time frame of the class period, and totaling between 311 and 621 instances of recorded mouse activity.

Some students' interaction frequencies were so similar, that their points on the graph overlap. Specifically, Participants 8, 9, and 11 exhibited interaction frequencies of 451, 483, and 454 respectively.

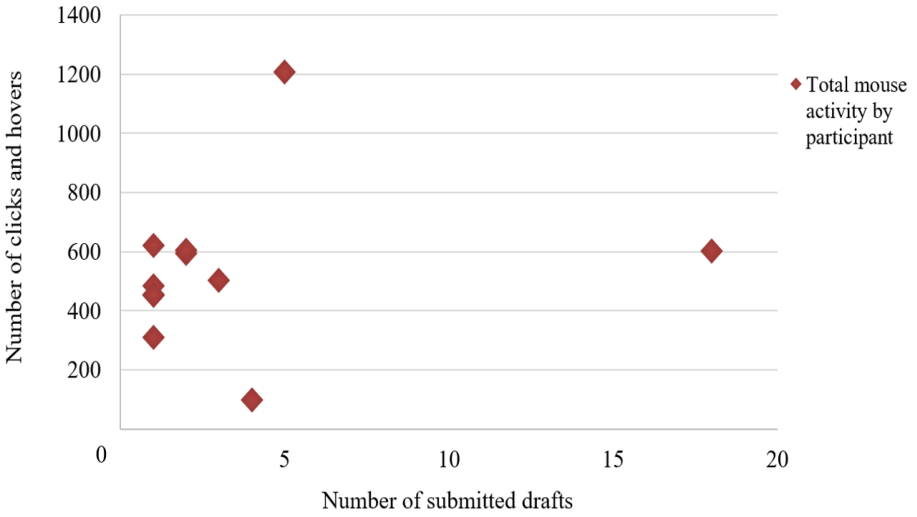

Figure 7. Frequency of mouse activity and draft submissions.

Our sequence analysis of the screen capture data further revealed how individual students progressed through their consumption of feedback and use of other features with a focus on the order of these interactions. In Figure 8, we attempt to depict the variation in individual participants' interactions with the specific features of the RWT (please note that Figure 7 above presents a more holistic interaction picture by relying on the total number of clicks and hovers on all the features). As it can be seen in Figure 8, where each color segment indicates individual students' interactions with a certain feature, each student proceeded in various ways. Five students started with the micro-level sentence-based feedback prompts on step functions (green), four – with macro-level feedback on moves (blue), and two – with the macro-level, graphical/numerical feedback (purple) and macro-level step feedback (orange). The overall sequences of their continued interactions varied considerably. What seems to be common, however, is that most participants' first text modifications in the editing box (yellow) were preceded by repeated access to multiple other features.

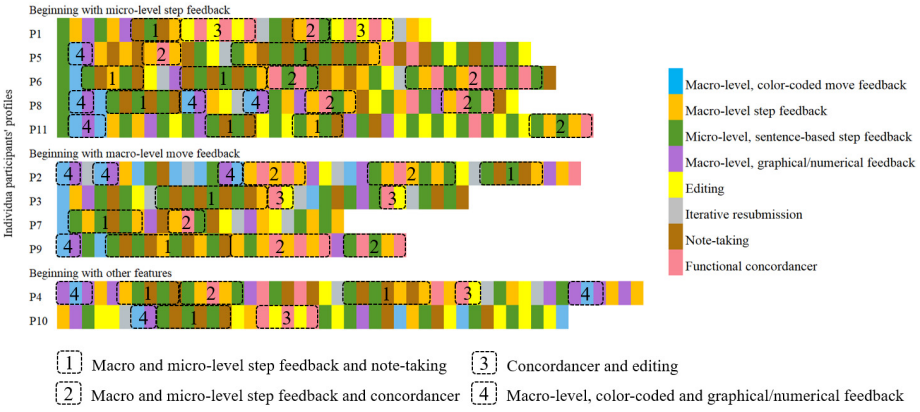

Figure 8. Sequences of individual interactions with the features of the Writing Research Tutor.

Albeit limited, this data reveals some tentative but interesting patterns of sequenced interactions, which are encircled and numbered in Figure 8. The first pattern was characterized by students' back-and-forth interaction with the macro- and/or micro-level feedback on steps (orange and green) and note-taking (in brown). The second pattern is similar, except that the students' use of the micro and macro-level feedback (green and/or orange) was accompanied by their access to the functional concordancer (pink). The third pattern highlights the connection between the uses of the concordancer (pink) and editing (yellow) features. The fourth pattern shows that two forms of the macro-level feedback on moves – color-coded (blue) and graphical/numerical (purple) were accessed in tandem. It was possible to observe this detail because these forms of feedback are displayed separately on the interface, and the screen capture videos allowed to see the respective mouse movements.

Although the students interacted with the macro- and micro-level types of feedback most frequently, the time-on-task analysis revealed that they spent a comparable amount of time on the less frequently accessed features. What can be observed in Figure 9, which illustrates a breakdown of the time participants spent interacting with features of the RWT, for example, is that students' processing of the sentences with step indicative prompts (micro-level feedback, part of pattern 1) accounted for 26% of their time, and 25% of their time was spent using the note-taking feature (also part of pattern 1). At 24%, the third greatest amount of time involved making modifications to the draft in the editing textbox (part of pattern 3). Interactions with other features that constituted less time included accessing examples from published RAs in the concordancer, the macro-level step feedback, and the color-coded and graphical/numerical types of macro-level move feedback (see Fig. 9).

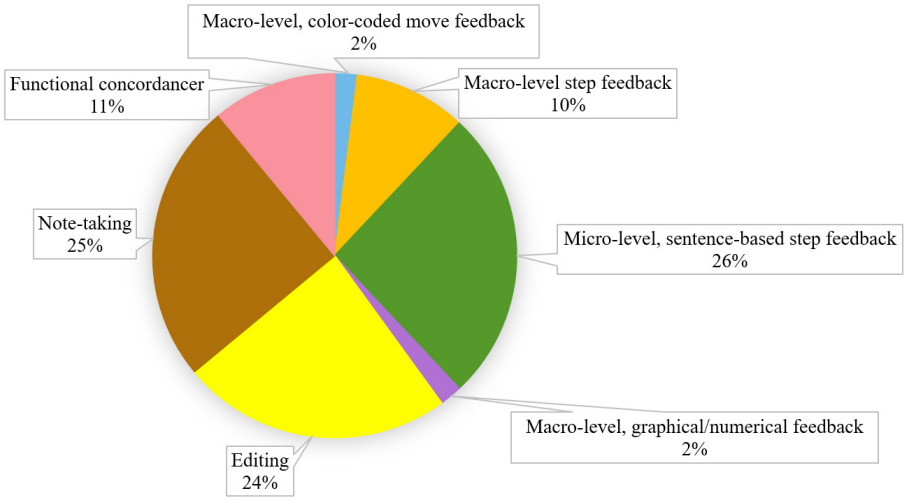

Figure 9. Participants' time-on-task with the features of the RWT.

5.2 Revision processes

To address the second research question focused on the impact of students' interactions with the RWT's features on their revision processes, the stimulated recall transcriptions were analyzed inductively and segmented into 1,726 idea units (151 on average per participant) pertaining to six emergent themes and respective sub-themes. These thematic categories suggest that interaction with the features seemed to have triggered fundamental metacognitive processes.

Table 3 is organized by prevalence of themes. The students tended to talk extensively about the *environment*, or the RWT's feedback and scaffolding (34% of idea units). *Strategizing*, or the selection, modification, and creation of strategies for revising represented text, was brought up by students in 21% of the data. *Reflection* of how they represented the rhetorical problem, plan, and standards of evaluation for the actual text occurred in 18% of the data. *Problem solving,* detecting and diagnosing problems in represented text, was also a relatively prominent theme (17%). Less prominent were *text production* for translating revisions from represented text to actual text (6%), and *text processing* – reading to represent and comprehend actual text (3%).

Table 3. Themes indicative of revision processes

Theme	Idea Units		Examples from stimulated recall data
Sub-theme	N	%	
Environment	587	34	"I think the software helped me identify the function." (Perception of helpfulness – P11)
Perception of helpfulness	271	46	
Awareness of features	119	20	
System personification	72	12	
External factors	60	10	
Affect	33	6	
Motivation	31	5	
Strategizing	370	21	"and then I try again to see if that works" (Taking local strategic actions – P4)
Taking local actions	319	86	
Changing approach	51	14	
Reflection	312	18	"and then I said, okay, this was right, this was right, this was right" (Self-evaluation – P2)
Self-evaluation	132	42	
Self-verification	130	42	
Drawing on knowledge	50	16	
Problem solving	299	17	"I can correspond the example sentence to my reading, writing" (Comparing representations – P9)
Detecting dissonance	166	56	
Diagnosing	100	33	
Comparing representations	33	11	
Text production	112	6	"and then I ended up saying that I collected data through multiple data measures" (Re-writing/ adding – P1)
Revising	87	78	
Re-writing/adding	21	19	
Resubmitting modifications	3	3	
Paraphrasing	1	1	
Text processing	46	3	"because I would try and go back and read careful what I had" (Critical reading – P6)
Critical reading	38	83	
Selecting input	1	2	
Noticing patterns	7	15	

The majority of responses about *environment* reflected the helpfulness of interacting with the RWT. The students indicated that they gradually developed an awareness of the different features and when to better use them. Worth mentioning is that students began to personify the tool. For example, one student said, "but it didn't make any sense for the system" (P1). Some students noted some factors that were not related to their hands-on use of the tool (e.g., "I wish I had more time to look at it" P9). There were also comments about motivation (e.g., "Yeah, but at the same time I still want to improve my writing" P4) and affect (e.g., "But then after I started liking this" P11). Such comments tended to occur when students mentioned having agreed with the sentence-level feedback by clicking on 'thumbs up.'

Disagreement with this type of feedback indicated by 'thumbs' down appeared to increase students' motivation to make more use of other features, as evidenced for instance by the second pattern of sequenced interactions in Figure 8.

For *strategizing*, taking local strategic actions (86%) included ignoring feedback, delaying, searching for examples of functional language in the concordancer, sequencing actions, note-taking, and acting upon the feedback and scaffolding. The RWT also seemed to help students change their approach to writing (14%) by helping them realize what works. Interestingly, one student imported a peer's paper into the tool and analyzed it, finding even that helpful. He said, "Yeah, I used the paper from a classmate to see how she did…she did a lot better than I thought… so it's kind of good for us to get this idea" (P2).

Two themes under *reflection* were self-evaluation (42%), which refers to the students' either positive or negative assessment and justification depending on whether they agree or disagree with the feedback, and self-verification (42%) as a way of verifying if intended functional meaning and sentence-based feedback were the same. Participant 1 demonstrated both these types of processes. She justified her agreement with the feedback and then verified her choice, "I agreed with the function of the system because I could also see that, yes, my sentence has this function but not as a primary goal". Other students were able to draw on general knowledge or knowledge of writing conventions (16%), as in the following student's response: "but sometimes we use parenthesis not for reviewing but for other purposes too" (P3).

Problem solving also emerged as a fundamental revision process, most often in the form of detecting dissonance (56%). When self-evaluation was negative, i.e. the student disagreed with the feedback, together with self-verification, it triggered students' noticing of a mismatch between intended and articulated meaning and how the RWT interpreted what they may have meant to express. For example, a student examined step-level feedback on a sentence (clicked on 'thumbs down' to disagree) and said, "I didn't have this in mind" (P7). The feedback prompted her to think and read more critically, as she questioned, "Why does the tool identify it as a different function?" Diagnosing (33%) was evident in students' revision processes when they would formulate a hypothesis about their writing: "And maybe those, if I figure out those, maybe that's the blind point that I have in my writing" (P4). The students compared the internal representation of their text (11%) mostly when thinking about what they meant to say with the step function detected by the tool.

Text production (6%), realized through reconsidering ways to revise the students' external representation of the text, making changes and additions, paraphrasing and resubmitting the modified text, was enabled by the availability of the editing textbox on the same page interface as the feedback, and by the "Analyze" button. *Text processing* (3%), when the students appeared to carefully read both their draft and the examples from the annotated texts in the RWT, notice

patterns of language use within a given step, and select language choices appropriate for their own rhetorical intent, was often scaffolded by the sentence-based step feedback and the concordancer. These two sub-themes occurred the least, although one may expect them to be more common. This was not the case in our study perhaps because the students used the tool for the first time and had a little more than one hour to revise with it. The time-on-task data showed that they spent more time on making sense of the sentence-level feedback likely trying to connect it with the internal representation of their text, a reflective process that appeared to precede problem solving. They also spent about the same amount of time on note-taking and editing, which may not necessarily lead to more text modifications within a restricted period of revising time.

Figure 10 graphs the themes indicative of revision processes per student in the same order as presented in Table 3, with the most frequent process (*environment*) shown first and the least frequent process (*text processing*) last. Here we see that the extent to which the students engaged in these processes varied, which resonates with the individual variation in interaction behavior with the RWT's features noted in the screen capture data (Figure 8). Nevertheless, in their stimulated recalls, most students referred to the *environment* and *strategizing* (roughly between 20% and 40%), which may denote that the former played an important role in fostering the latter. These two themes surface to a similar extent even for the two outliers who stood out due to the far more frequent mouse activity (P2) and number of submissions (P4). Interestingly, P11 talked about *strategizing* more than *environment;* yet, this participant appeared to take local strategic actions

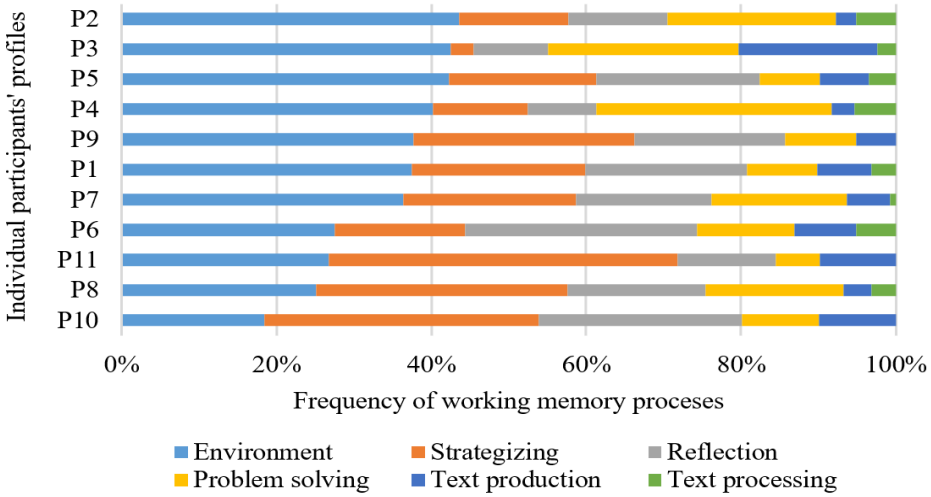

Figure 10. Themes indicative of revision processes per participant.

due to the interaction within the RWT environment. In fact, early engagement with the environment seemed to stimulate more strategic processes later as the students continued to revise with the tool.

5.3 Text modifications

To determine whether and how the students' interaction with the RWT's features impacted how they modified their text during the execution process, we first coded each sentence in their initial and last drafts for a move and a step.

In total, the first drafts contained 252 sentences coded for move 1, 11 sentences coded for move 2, 25 sentences coded for move 3, and the last drafts contained 249, 22, and 36 move-coded sentences, respectively. Then, we compared the number of move/step occurrences from first to last drafts to see whether the students made move-step related revisions and how they were done. Wilcoxon Signed-Ranks test results showed that the move/step occurrences in the last draft (Mdn = 4) were significantly higher than in the first draft (Mdn = 2; z = -2.653, $p = .008$, $r = .64$).

Table 4 renders the within-student comparison based on the Wilcoxon Signed-Ranks Test for individual differences in rhetorical composition. Statistically significant results are marked with asterisks for seven out of eleven students. Interestingly, the within-groups Wilcoxon Signed-Ranks Test comparisons showed that the move/step occurrences in the nonnative speakers' (NNSs) final drafts (Mdn = 3) were significantly higher than in their first drafts (Mdn = 2; z = -2.046, $p = .041$, $r = .50$), whereas the native speakers' (NSs) final drafts (Mdn = 1) were not significantly different than their first drafts (Mdn = 0; z = -1.633, $p = .102$, $r = .40$).

Table 4. Wilcoxon Signed-Ranks Test for individual differences in rhetorical composition

Student ID	z	p	r
P1_NNS	-2.56	.011*	.57
P2_NNS	-2.30	.022*	.51
P3_NNS	-1.38	.168	.31
P4_NNS	-2.06	.040*	.46
P5_NNS	-1.99	.047*	.44
P6_NS	-2.61	.009**	.58
P7_NS	-.48	.630	.12
P8_NS	-.80	.423	.18
P9_NNS	-2.41	.016*	.54
P10_NS	-2.82	.005**	.63
P11_NNS	-.90	.369	.20

Note. * significant at p < .05, ** significant at p < .01

This may serve as indirect evidence in support of explicit genre-based pedagogy for language learners, advocated for in the field of English for Specific Purposes.

Figure 11 further differentiates between students' first and last drafts that show modifications in the move/step composition that can be considered significant and not significant. Overall, all the students used move 1 most extensively in first and final drafts. The two steps of move 1 that are most prominent were 'providing general background' and 'reviewing previous research.' For drafts that fell in the category of significant modifications, substantial changes are noticeable in the distribution of these two steps, with students especially expanding 'reviewing previous research.' Although 'providing general background' was used to a lesser extent in the last drafts, this does not mean that the participants simply deleted text from their first drafts. On the contrary, deletions were rare, as the students tended to support ideas with references often modifying the content and language choices and transforming or integrating the content of earlier sentences into 'reviewing previous research.' As for other steps, they had more of 'introducing present research purposefully' and 'presenting research questions,' and added 'claiming centrality,' 'indicating a gap,' 'highlighting a problem,' 'stating the value of present research,' and 'outlining the structure of the paper.'

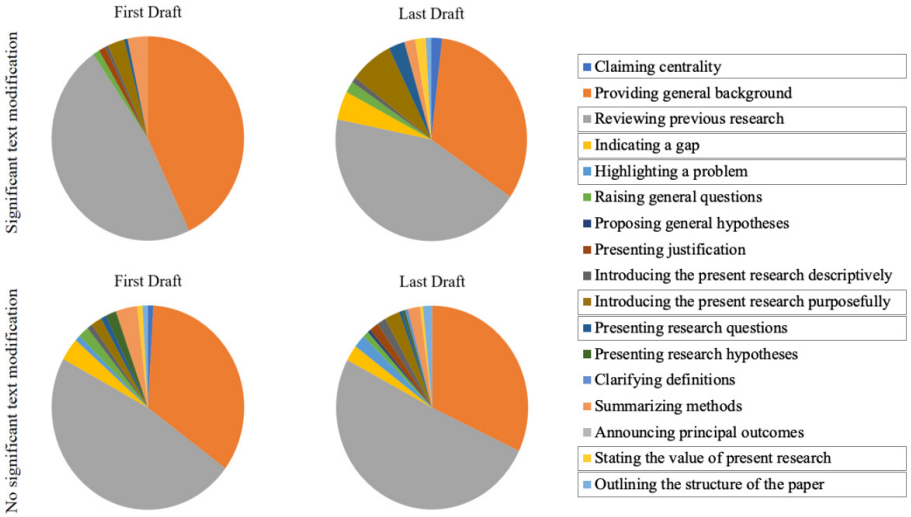

Figure 11. Significant and non-significant move/step modifications from first to last drafts.

The pie charts representing the drafts that were not significantly modified indicate minor changes in the use of steps. The first drafts already contained a wider range of steps, which may suggest that the authors of those initial texts started more strongly in terms of applying the rhetorical conventions they studied in the course. In the last drafts, however, they did not expand on the previously included steps, and only a few students added new steps such as 'proposing general hypotheses' and 'presenting a justification.' Moreover, unlike the charts at the top of Figure 11, the charts at the bottom show almost no difference in the distribution of 'providing general background' and 'reviewing previous research,' but their amount is comparable with the drafts in the other category.

At this point, it is worth comparing Figure 11 with Appendix B, where similar pie charts represent the use of steps in the Introductions from the published articles in the RWT corpus, including Introductions in the participants' disciplines. Despite the fact that the students had little time to revise in class, their last drafts somewhat approximate the step distribution in high quality texts of this genre. 'Reviewing previous research' is the biggest step in both cases, and all the last drafts begin to exhibit the variety of steps in published texts. It is very possible that, had the students had more time to revise with the RWT, they would have increased their use of different steps.

One may infer that students were on the path of crafting their texts by integrating "a constellation of recognizable forms bound together by an internal dynamic" (Campbell & Jamieson, 1978, p. 21), which expert writers use to produce the rhetorical effect expected of an article Introduction.

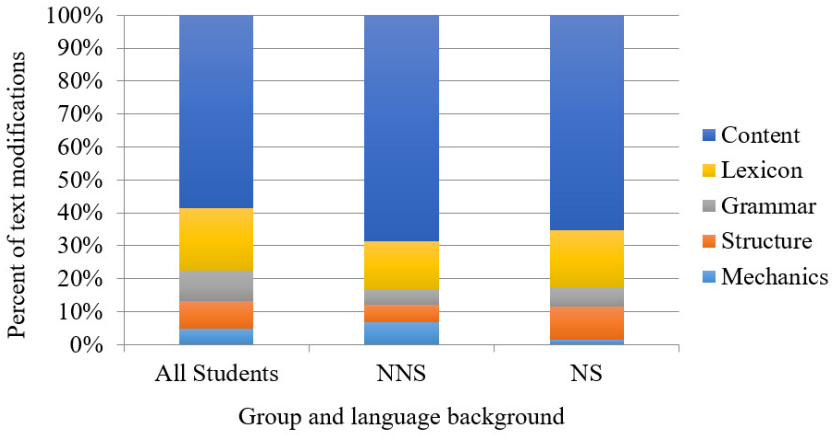

Figure 12. Text modifications for group and language background.

While the number and distribution of the steps contribute to the quality of published articles, high quality also depends on the writer's ability to compose multi-functional discourse by integrating the functional meanings of more than one step in a sentence or even a clause (Cotos, 2019; Cotos et al., 2015; Cotos, Huffman, & Link, 2017b). This is what we particularly noticed in the last drafts with significant modifications.

The actual text modifications were examined by mapping students' first drafts onto their last drafts. This comparison revealed the five types of text modifications presented in Figure 12 (above): content (e.g., additions, deletions, modified ideas); lexis (e.g., move-specific, non-move-specific); grammar (e.g., verb tense or form, subject-verb agreement, plurals), structure (e.g., paragraph, syntactic structure), and mechanics (e.g., punctuation, citation format).

Table 6 complements Figure 11 with examples of these types of text modifications.[4] Changes that the students made from first to last draft are in italics. For instance, the revision to content points to the students' increased awareness that more evidence is needed to support the claim originally made in the first draft. The revision of lexicon replaced the introductory words (along with a change in grammar) to emphasize a contrast of ideas leading to indicating a gap. The grammatical change was a correction from plural to singular form of a noun phrase. The fourth example shows a structural shift between the main and secondary clauses, and the last one places the parenthetical citation inside the sentence before the comma.

Table 6. Text modifications and examples

Text modifications	First draft	Last draft
Content P8_NS	Despite the solids-handling industry being quite mature, the design and operating conditions of mixing equipment remains open to speculation and lacks quantitative justification.	Despite the solids-handling industry being quite mature, process and equipment design and the selection of operating conditions remains open to speculation because it lacks quantitative justification, *being based on judgment and trial-and-error procedures, rather than quantitative science [6].*
Lexis P1_NNS	Nevertheless, there are studies analyzing the effects of keeping classmates peers together in elementary school.	*Despite numerous* studies on peer effects in general and on education in particular, the study of the effects of keeping classmates peers together during elementary school *has been neglected.*

Grammar P8_NS	An improved sampling procedure is needed thus providing a solution to these common encountered problems.	An improved sampling procedure is needed thus providing a solution to *this commonly encountered problem.*
Structure P11_NNS	While DHA plays an important role in development of the nervous system, including vision, during fetal and infant development, there is evidence of continued beneficial effect on brain health from DHA beyond fetal development.	In addition to the important role of DHA in development of the nervous system during fetal and infant development, there is evidence of continued beneficial effect on brain health from DHA beyond fetal development.
Mechanics P3_NNS	For example, more than one surface plasmon mode can be excited due to the opportunity for coupling to several diffracted orders from the grating interface.[13]	For example, more than one surface plasmon mode can be excited due to the opportunity for coupling to several diffracted orders from the grating interface *[13].*

6. Discussion

Collectively, the results suggest that interactive behaviors in a writing environment altered by technology such as the RWT, with automated rhetorical feedback and scaffolding, has the potential to foster metacognitive processing during revision. Before we discuss the findings with regards to each research question, let's consider a scenario derived from our data which describes a student's experience engaging with this tool. (Amy is this students' pseudonym.)

> Amy is a graduate student in Agricultural and Bio-Systems Engineering who just drafted a research paper. She logs in to the RWT, chooses her discipline, copy-pastes her draft of the Introduction section into the tool, and clicks on "Analyze." Within seconds, her draft comes back color-coded for moves, and she can see the rhetorical structure of her argument. The macro-level, graphical/numerical feedback at the right tells her whether her draft is similar to Introductions written by experts in her discipline. The bar graphs show that two of her moves fall outside the goal range compared to texts published in her field. She wonders, "Where do I fall short?" and clicks below the graph to see which rhetorical steps in her "identifying a niche" move need work. She gets excited to see feedback saying that she did a good job on 'indicating a gap' and 'highlighting a problem.' She is also motivated

> to improve because the feedback tells her what her draft is lacking. She notices that the tool suggests that her draft is lacking 'presenting a justification,' but she thought she did write about why her research is important. She goes back to her color-coded draft and finds the sentence where she thinks she is justifying her research: "Our contribution is to develop a measurement technique capable of determining the mixing effectiveness of a laboratory-scale double screw cold-flow pyrolyzer." The feedback for this sentence hints, "You are likely stating the value of your study." "How is that different from presenting a justification?" wonders Amy, and hovers over Learn More to see the definitions of these two steps she doesn't seem to understand very well. This shortcut helps her realize that she, indeed, means to emphasize the importance of her study and is thus making a claim about how it addresses the niche, rather than arguing for the need to fill the gap she identified in previous research. She decides to keep this sentence, but also goes back to the text to add, "Effective laboratory-scale cold-flow pyrolyzers are important." When she re-analyzes her draft, the feedback on this new sentence asks, "Are you providing general background?" Now, while she knows what she means, her meaning is apparently not yet clearly expressed. Amy clicks on "Examples" to see how expert writers express similar claims. Here, she sees many sentences that the RWT's concordancer has extracted for her from the corpus in her discipline. All the sentences present a justification, but each sentence does it in its own way. She notices the kind of language authors use in these types of claims, goes back to her sentence, successfully changes it to "Hence, an improved procedure is needed to provide a solution to this commonly encountered problem," and moves on to other issues that the tool helps her detect.

Throughout this scenario, Amy continues interacting with the feedback and scaffolding features of the RWT when revising her Introduction draft. The nature of this interaction was the focus of our RQ1. Analyses of interaction logs and screen captures showed a great deal of variation in the way students accessed the features. Nevertheless, they interacted most frequently with the macro-level move/step feedback and the micro-level sentence-based step feedback, and these features were present in three of the four sequential patterns of interaction that surfaced in the individual participants' profiles (see Figure 8). Although the patterns we delineated are tentative, for they were deduced from data gathered during only one class period, their meaning can still be interpreted in relation to theory and more specifically to our augmented cognitive writing model (Figure 5).

The first pattern, where the sequence of students' actions included macro- and micro-level feedback on steps and note-taking, could be interpreted as students' close reading of the feedback to better detect or diagnose a particular problem in the represented text, which is theorized as a key aspect of problem representation.

The second pattern, where students accessed the micro and macro-level feedback along with the concordancer, may reflect the evaluation process, as the students read representative examples from the corpus attempting to better understand how the rhetorical problem can be more effectively addressed by using specific functional language choices. Pattern three, linking the use of the concordancer and editing, suggests that the students may have been developing a strategy, i.e. to determine linguistic instantiations of the steps and then proceed to execution by incorporating appropriate language when modifying their text. Students' use of the two forms of the macro-level feedback on moves, color-coded and graphical/numerical, which makes the fourth pattern, typically occurred either at the beginning or later during the interaction with other features. When these forms of feedback were accessed early, the students juxtaposed the overall distribution of the moves in their drafts and in published Introductions from their disciplinary corpus, and that may be indicative of problem representation. Later on, when checking the color codes and graphical/numerical feedback while interacting with other features, the students were trying to better comprehend how the steps may contribute to the composition of the moves, this possibly being a sign of the evaluation process.

These theoretical interpretations are supported by the themes that emerged in the stimulated recall data, which was collected to shed introspective light on RQ2 – whether and how the interactional behaviors with the RWT's feedback and scaffolding features may impact their revision processes. Five themes (see Table 3) could be mapped onto deliberate working memory processes described by Butterfield et al. (1996) and other cognitive writing models. A sixth theme, *environment* (34% of idea units), was present in the data as an external factor, which we introduced into the augmented revision model (see Figure 5) by integrating multi-level automated feedback and scaffolding. The *environment* was perceived as being helpful and offering a motivational experience.

From students' explicit mentions as well as from our observations of the screen captures, some connections between the themes indicative of metacognitive processes and specific features can be made. The idea units coded as *strategizing* (21%) generally referred to the concordancer and note-taking features. *Reflection* (18%) seemed to be commonly prompted by the sentence-based step feedback and consolidated through the use of the interactive 'thumb' markers. *Problem solving* (17%) was triggered by the color-coded and comparative graphical/numerical forms of the move feedback and by the sentence-based step feedback – all helping the students detect and diagnose problems in the represented text. The graphical/numerical form of the feedback also seemed to have played a motivational role because many students wanted to revise until their draft approximated the average distribution of the moves/steps in their discipline. *Text processing* (3%) was often scaffolded by the sentence-based step feedback and the

concordancer. Finally, *text production* (6%) was supported by the editing textbox and the "Analyze" button for iterative resubmission. With regards to individual students' profiles based on the frequency of these revision processes, there was a certain degree of variation (see Figure 10), which is similar to individual interactions with the RWT's features (Figure 8). Possibly, variation in interactional behaviors may have influenced which revision processes were triggered and the extent to which they may have been fostered.

The interaction with the RWT's feedback and scaffolding features also had an impact on the actual text modifications the students made. This finding addresses our RQ3. Statistical and comparative analyses of students' first and last drafts showed that the students modified their texts both globally and locally. Globally, they made substantial changes to the rhetorical composition of their drafts. Overall, the students' application of move/step modifications suggests that the effects of their interaction with the multi-level feedback and scaffolding began to positively transfer to producing rhetorically appropriate writing. This is in line with the evidence provided by Knight et al. (2020) whose students showed text improvement by incorporating rhetorical moves after having received feedback from the AcaWriter.

While it cannot be confidently claimed that similar revisions would or would not have occurred if revising without the RWT, inferences can still be made drawing on the students' use of its features. The last drafts with a significantly modified rhetorical composition proportionately approximated the distribution of moves/steps in the disciplinary texts from the corpus (Appendix B), which can be linked to the goal-orienting and comparative nature of the graphical/numerical form of the macro-level feedback, as well as to the on-demand availability of the concordancer. Additionally, based on our observations as we analyzed the screen capture data, move/step-specific changes or additions stemmed from interconnected interactions with different features of the tool. For instance, Participant 5's addition of the step 'introducing present research purposefully' was prompted by:

a) noticing that he had very little move 3 (green color in his draft) → macro-level, color-coded move feedback;
b) looking at the range bars and seeing that it was far from the average distribution of move 3 in his discipline → macro-level, graphical/numerical feedback;
c) clicking to expand the move 3 range bar and seeing that the draft was lacking the step 'introducing present research purposefully' → macro-level step feedback;
d) writing that he needs to include it → note-taking;
[repeating actions (c) and (d) noting that he needs to work on another step of move 3 as well]

e) clicking on 'Examples' and scrolling down to read a number of sentences representative of the step he needed to better understand → concordancer.

Such interconnected interactions with the RWT's feedback and scaffolding features could be clearly seen in the sequential patterns. On average, each student interlaced three patterns. The students whose last drafts showed significant changes, and thus improvement in the rhetorical composition of their Introductions, engaged in three or four types of sequential patterns.

The move/step revisions were in line with the rhetorical conventions of the Introduction part-genre, and this is not unexpected because the RWT operationalized these particular conventions. Importantly, the last drafts contained other types of genre appropriate text modifications. Prevalent were content changes, adding or modifying ideas needed to accomplish communicative goals. Second to that were modifications in step-related lexicon, or functional language that was needed to more explicitly convey rhetorical intent. Interestingly, grammar, syntax, and mechanics were also edited. Such modifications are expected when using automated evaluation tools that provide feedback on these aspects of writing (e.g., Criterion, Grammarly). It is thus encouraging to see that automated rhetorical feedback can indirectly support *problem solving* and *text production* processes relative to local aspects of writing. Finally, although we did not specifically investigate the connection between students' level of agreement with micro-level feedback (captured through the 'thumbs' feature) and text modifications, we must note that the screen capture and stimulated recall data suggest that disagreement seemed to actually be beneficial, as it prompted the text modifications focused on rhetorical composition.

7. Indications for future developments

The constructivist concepts of feedback and scaffolding (Vygotsky, 1962) have been widely applied to promote better learning with technology (Sawyer, 2006), for computer-enhanced scaffolding has full potential to transform learning. Scaffolding should address metacognitive activities to enable students to correct misconceptions while also integrating new ideas in conceptual understandings for problem solving (Kim & Hannafin, 2011). From this perspective, we will pursue future design and research of new features for the RWT that would further leverage learning theory as well as advanced computational methods. Our envisioned developments will align with two key areas for potential digital writing tools development recommended by Strobl et al. (2019): "stronger integration of strategy instruction, ideally linked to feedback" and "stronger focus on macro-level feedback focusing on the writing goals and genres" (p. 45).

7.1 Guiding revision with strategic scaffolding features

In future development, we envision complementing the features of the RWT with new strategic scaffolding drawing on existing guidelines for the design of metacognitive scaffolding systems (Roll, Aleven, McLaren, & Koedinger, 2007). We expect that providing strategic scaffolding directing students towards more effective uses of the features will help them acquire the ability to efficiently and frequently activate and monitor cognitive and metacognitive processes. The strategic scaffolding would include guidance for help-seeking within the RWT and guidance for metacognitive strategies. For support with help-seeking, we would delegate some of the responsibility to the system, rather than to the students (Luckin & du Boulay, 1999), to stimulate their noticing of problems and recognize the need for help.

The RWT system would track students' usage of rhetorical feedback and conceptual scaffolding resources that can lead to the resolution of writing problems (e.g., accessing examples in the concordancer, which students may review immediately before resolving the issue of lacking a step in a move). The guidance for metacognitive strategies would target metarhetorical (awareness of one's self as a writer), metastrategic (knowing what and when a strategy works for one's self and when it does not) and metalinguistic (knowing what language better expresses the intended meanings) awareness (Horning, 2002). For example, sentence-based prompts like "You have changed the words in this sentence several times. How about [this form of conceptual scaffolding]?" would aim to prompt strategy selection for effective and appropriate text modification.

7.2 Guiding text modifications with context-aware rhetorical feedback

We would also strive to develop additional feedback at the whole-text level in order to direct the students toward specific disciplinary patterns in terms of how to unfold parts of their research story in a given section of the article. As described in Cotos et al. (2017b), the order of steps can be conceived as the "DNA" of the section, and it can be computationally detected and modeled. A recent study on context-aware computational models for automated analysis of rhetorical structure of the research article (Fiacco, Cotos, & Rose, 2019) can inform the design of such feedback based on the expected progression of moves and steps in a given discipline compared to the use of moves and steps by a student. The feedback may for instance suggest, "The writers in [student's discipline] tend to open their Introduction by claiming centrality of their topic," or that they "tend to justify the need to conduct research immediately after they indicate a gap." This type of context-aware feedback could prompt the students not just what step-related modification to make, but where it should be made.

8. Conclusion

The results of this study allow for the following deductions. Overall, the variation in students' interactions with the RWT indicates that they behaved autonomously, individualizing their own paths in a dynamic revision process with this tool. Varied interactions with different types and forms of feedback may foster students' close and deliberate examination of their produced text, and their integration of functional language that is indicative of specific rhetorical steps, as evidenced by students' use of the concordancer in tandem with note-taking and editing. This combination of actions may help the students grasp and represent the rhetorical problem, detect dissonances in their draft, plan and effectively execute appropriate text modifications against a backdrop of authentic genre-relevant examples produced by published authors in their disciplines.

Our findings are valuable in that theoreticians, teachers, and tool developers can rely on this empirical evidence to make decisions about the uses of genre-based tools like the RWT. More work remains to be done still, for we only provide a snapshot obtained during a limited time with a small number of participants. Theoretically, we are interested in understanding how the RWT may help student writers move from knowledge telling towards knowledge transformation (Bereiter & Scardamalia, 1983) and transition from peripheral to full participation in the discourse of their disciplinary communities. Following up on this study, where there was not sufficient data for identifying interaction sequence patterns statistically (as opposed to making inferences from their occurrence in most individual students' interactions with the features of the tool), it would be interesting to examine such patterns based on their receptive versus productive nature. In other words, would there be any patterns around what receptive features (e.g., feedback, concordancer) may lead to the use of the productive features (e.g., notetaking, editing, iterative resubmission)? Furthermore, much more could be learned about text processing and text production, the least frequent processes we identified, if the analysis of students' texts expanded to examining all the incremental revisions between the first and last drafts.

In terms of practice, future research could measure different aspects of writing quality to determine how the draft quality may affect the use of the tool's features and vice-versa. New studies could also conduct a deeper examination of how the overall quality of texts improves when using the RWT as compared with a tool like Grammarly, and how those improvements may differ when no such digital writing environments are used. Additionally, the course where the RWT was implemented in this study and other contexts where it has found more uses (e.g., research writing workshops, individual consultations with writing tutors, peer review groups) aim to help students produce publishable manuscripts. We plan to examine longitudinal data gathered in these contexts, as teachers and faculty advisers would appreciate

knowing whether and how this tool may improve the chances of students' manuscripts becoming published.

The landscape is fertile for bridging theory and practice through the use of digital writing technologies. Genre-based automated writing evaluation is a nascent domain with promising potential for various forms of feedback and scaffolding designed to promote writing development. We look forward to the appearance of platforms similar to the RWT for other academic and professional genres – theoretically grounded, empirically validated, and practically effective.

Notes

1. In research writing, the task is to present a scientific argument to a target disciplinary community.
2. These stages are also known as cognitive, associative, and autonomous (Fitts & Posner, 1967) or declarative, procedural, and automatic (Anderson, 1983).
3. The Cohen's kappa coefficient (κ) for the human-RWT reliability was .84 for moves and .71 for steps (also see Cotos et al., 2017b).
4. The examples in Table 5 were selected to represent most of the coding, as each type of text modification was generally clearly distinguishable. However, there were instances when a text modification was coded as more than one type, and this was done more for grammar and structure, as well as for mechanics (e.g., changes in punctuation along with a change in syntax).

Acknowledgements

The authors express their appreciation to the anonymous reviewers and to the editors for their thorough and insightful suggestions. Special thanks to the research assistant Elizabeth Lee and to all the students who participated in this study.

References

Abel, S., Kitto, K., Knight, S., & Buckingham Shum, S. (2018). Designing personalised, automated feedback to develop students' research writing skills. In *35th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education*. Melbourne, Australia. Retrieved from http://sophieabel.net/wp-content/uploads/2019/02/Improving-research-students-writing-with-writing-analytics.pdf

Allal, L, Chanquoy, L., & Largy, P. (2004). *Studies in Writing: Vol. 13, Revision. Cognitive and Instructional Processes*. Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-007-1048-1

Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2016). Computer-based writing instruction. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.). *Handbook of writing research* (pp. 316–329). (2nd ed.). New York, NY: Guilford. https://doi.org/10.1007/s11145-008-9121-2

Anderson, J.R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Badley, G. (2009). Academic writing as shaping and re-shaping. *Teaching in Higher Education 14*(2), 209-219. Doi https://doi.org/10.1080/13562510902757294

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum Associates. https://doi.org/10.2307/327270

Benetos, K., & Bétrancourt, M. (2020). Digital authoring support for argumentative writing: What does it change? *Journal of Writing Research, 12*(1), 263-290. https://doi.org/10.17239/jowr-2020.12.01.09

Bransford, J. D., Brown, A. L., & Cocking, R.R. (2000). *How People Learn: Brain, Mind, Experience, and School.* Washington: National Academies Press. https://doi.org/10.17226/9853

Butterfield, E., Hacker, D., & Albertson, L. (1996). Environmental, cognitive, and metacognitive influences on text revision: Assessing the evidence. *Educational Psychology Review, 8*(3), 239-297. https://doi.org/10.1007/bf01464075

Byrne, D. (1986). *Teaching Oral English* (2nd ed.). Harlow, UK: Longman.

Campbell, K. K., & Jamieson, K. H. (1978). *Form and genre: Shaping rhetorical action*. Falls Church, VA: Speech Communication Association.

Chanquoy, L. (2009). Revisions Processes. In: R. Beard, D. Myhill, J. Riley & M. Nystrand (Eds.) *The Sage Handbook of Writing Development* (pp. 80-97). London, Sage. https://doi.org/10.4135/9780857021069.n6

Chenoweth, A., & Hayes, J. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication, 18*, 80-98. https://doi.org/10.1177/0741088301018001004

Cotos, E. (2009). Designing an intelligent discourse evaluation tool: Theoretical, empirical, and technological considerations. In C. A. Chapelle, H.- S. Jun, & I. Katz (Eds.), *Developing and Evaluating Language Learning Materials* (pp. 103–127). Ames, IA.

Cotos, E. (2011). Potential of automated writing evaluation feedback. *CALICO Journal, 28*(2), 420-459. https://doi.org/10.11139/cj.28.2.420-459

Cotos, E. (2012). Towards effective integration and positive impact of automated writing evaluation in L2 Writing. In G. Kessler, A. Oskoz & I. Elola (Eds.), *Technology across writing contexts and tasks*, CALICO Monograph Series (Vol. 10, pp. 81–112). CALICO: San Marcos, TX.

Cotos, E. (2014). *Genre-based automated writing evaluation for L2 research writing: From design to evaluation and enhancement*. Basingstoke, UK: Palgrave Macmillan. https://doi.org/10.1057/9781137333377

Cotos, E. (2016). Computer-assisted research writing in the disciplines. In S. A. Crossley & D. S. McNamara (Eds.), *Adaptive educational technologies for literacy instruction* (pp. 225-242). NY: Taylor & Francis, Routledge. https://doi.org/10.4324/9781315647500-15

Cotos, E. & Huffman, S. (2013). Learner fit in scaling up automated writing evaluation. *International Journal of Computer-Assisted Language Learning and Teaching, 3*(3), 77–98. https://doi.org/10.4018/ijcallt.2013070105

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. https://doi.org/10.1016/j.jeap.2015.05.004

Cotos, E., Huffman, S., & Link, S. (2017b). A Move/Step model for Methods sections: Demonstrating rigour and credibility. *English for Specific Purposes, 46*, 90-106. https://doi.org/10.1016/j.esp.2017.01.001

Cotos, E., Link, S., & Huffman, S. (2017a). Effects of DDL technology on genre learning. *Language Learning and Technology, 21*(3), 104-130.

Cotos, E., & Pendar, N. (2016). Discourse classification into rhetorical functions for AWE feedback. *CALICO Journal, 33*(1), 92-116. https://doi.org/10.1558/cj.v33i1.27047

Cresswell, J., & Clark, V. (2011). Designing and conducting mixed methods research (2nd ed.). Thousand Oaks, CA: Sage Publications.

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing, 18*(1), 7–24. https://doi.org/10.1016/j.asw.2012.10.002

Deane, P., Quinlan, T., Odendahl, N., Welsh, C., & Bivens-Tatum, J. (2008). *Cognitive models of writing: Writing proficiency as a complex integrated skill. CBAL literature review writing* (ETS Research Report No. RR-08-55). Princeton, NJ: ETS. https://doi.org/10.1002/j.2333-8504.2008.tb02141.x

DeKeyser, R. M. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511667275

Edgar, R. C., & Batzoglou, S. (2006). Multiple sequence alignment. *Current opinion instructural biology*, *16*(3), 368–373. https://doi.org/10.1016/j.sbi.2006.04.004

Fitts, P.M., & Posner, M.I. (1967). *Human performance.* Belmont, CA: Brooks/Cole.

Fitzgerald. J. (1987). Research on revision in writing. *Review of Educational Research, 57*(4), 481-506. https://doi.org/10.3102/00346543057004481

Flower, L., & Hayes, J.R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365-87. https://doi.org/10.2307/356600

Flower, L., Hayes, J., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, Diagnosis, and the Strategies of Revision. College Composition and Communication. 37. 16-55. https://doi.org/10.2307/357381

Gass, S., & Mackey, A. (2006). Input, interaction and output: An overview. *AILA Review, 19,* 3-17. https://doi.org/10.1075/aila.19.03gas

Gass, S., & Mackey, A. (2007). Input, interaction, and output in second language acquisition. In J. Williams & B. VanPatten (Eds.), Theories in second language acquisition: An introduction (pp. 175-199). Mahwah, NJ: Lawrence Erlbaum. https://doi.org/10.1017/s0272263108080418

Galbraith, D. (2009). Writing as discovery. In *BJEP Monograph Series II, Number 6-Teaching and Learning Writing, 5*(26), 5-26. https://doi.org/10.1348/978185409X421129

Giddens, A. (1979). *Central Problems in Social Theory*. Berkeley, CA: University of California Press. https://doi.org/10.1177/000271628044800153

Hacker, D. J., Keener, M. C., & Kircher, J. C. (2009). Writing is applied metacognition. In D. J. Hacker, J. Dunlosky, A.-C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 154–172). New York, NY: Routledge. https://doi.org/10.4324/9780203876428

Halliday, M. A. K., & Hasan, R. (1989). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Oxford: Oxford University Press. https://doi.org/10.2307/3586740

Hayes, J. R. (2004). What triggers revision? In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision: Cognitive and instructional processes* (pp. 9-20). Dordrecht, Germany: Kluwer Academic. https://doi.org/10.1007/978-94-007-1048-1_2

Hayes, J. R. (2012). My past and present as writing researcher and thoughts about the future of writing research. In V. Berninger (Ed.), *Past, present, and future contributions of cognitive writing research to cognitive psychology* (pp. 3–26). New York: Psychology Press. https://doi.org/10.4324/9780203805312

Hayes, J. R., Flower, L., Schriver, K. A., Stratman, J. F., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics* (Vol. 2, pp. 176-241). New York, NY: Cambridge University Press. https://doi.org/10.1017/cbo9780511571374

Heift, T. (2002). Learner control and error correction in ICALL: Browsers, peekers and adamants. CALICO, 19(3), 295-313. https://doi.org/10.1558/cj.v19i2.295-313

Horning, A. (2002). *Revision Revisited*. Cresskill, NJ: Hampton Press.

Hoyer, P.O. (2004). Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research, 5*, 1457–1469.

Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing, 12*, 17-29. https://doi.org/10.1016/s1060-3743(02)00124-8

Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing, 16*, 148-164. https://doi.org/10.1016/j.jslw.2007.07.005

Jackson, S.L., Krajcik, J., & Soloway, E. (1998). The design of guided learner-adaptable scaffolding in interactive learning environments. *Proceedings of SIGCHI conference on Human Factors in Computing Systems 1998*. (pp. 187-194.) Los Angeles, CA. https://doi.org/10.1145/274644.274672

Jolliffe, D. A., & Brier, E. M. (1988). Studying writers' knowledge in academic disciplines. In D. A. Jolliffe (Ed.), *Advances in writing research: Vol. 2. Writing in academic disciplines* (pp. 35-77). Norwood, NJ: Ablex Publishing Company.

Kim, M. C., & Hannafin, M. J. (2011). Scaffolding problem solving in technology-enhanced learning environments (TELEs): Bridging research and theory with practice. *Computers & Education, 56*(2), 403–417. https://doi.org/10.1016/j.compedu.2010.08.024

Knight, S., Shibani, A., Abel, S., Gibson, Ryan, P., Sutton, N., Wight, R., Lucas, S., Sándor, Á., Kitto, K., Liu, M., Vijay Mogarkar, R., & Buckingham-Shum, S.J. (2020). AcaWriter: A learning analytics tool for formative feedback on academic writing. *Journal of Writing Research, 12*(1), 141-186. https://doi.org/10.17239/jowr-2020.12.01.06

Kroll, B. (1977). Combining ideas in written and spoken English: A look at subordination and coordination. In E. Ochs & T. Bennett (Eds.), *Discourse across time and space* (Southern California Occasional Papers in Linguistics, 5). Los Angeles, University of Southern California.

Luckin, R., & du Boulay, B. (1999). Ecolab: The development and evaluation of a Vygotskian design framework. *International Journal of Artificial Intelligence in Education, 10*(2): 198–220. https://doi.org/10.1007/s40593-015-0072-x

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics, 27*(3), 405-430. https://doi.org/10.1093/applin/ami051

Martin, J. R. (1985). Process and text. In J.D. Benson & W.S. Greaves (Eds.), *Systemic perspectives on discourse* (Vol. 1, pp. 248-274). Norwood, NJ: Ablex Publishing. https://doi.org/10.5040/9781474211932.ch-008

Martin, J. R. (1993). A contextual theory of language. In B. Cope, & M. Kalantzis (Eds.), *The powers of literacy. A genre approach to teaching writing*. London: The Falmer Press. https://doi.org/10.4324/9780203149812

Mizumoto, A., Hamatani, S., & Imao, Y. (2017). Applying the bundle–move connection approach to the development of an online writing support tool for research articles. *Language Learning, 67*(4), 885–921. https://doi.org/10.1111/lang.12250

McCutchen, D., Teske, P., & Bankston, C. (2008). Writing and cognition: Implications of the cognitive architecture for learning to write and writing to learn. In C. Bazerman (Ed.), *Handbook of research on writing* (pp. 447–465). Mahwah, NJ: Erlbaum.

Nix, I., & Wylie, A. (2011). Exploring design features to enhance computer-based assessment: Learners' views on using a confidence-indicator tool and computer-based feedback. British Journal of Educational Technology, 42(1), 101-112. https://doi.org/10.1111/j.1467-8535.2009.00992.x

Palermo, C., & Wilson, J. (2020). Implementing automated writing evaluation in different instructional contexts: A mixed-methods study. *Journal of Writing Research, 12*(1), 63-108. https://doi.org/10.17239/jowr-2020.12.01.04

Pallant, J. (2007). SPSS survival manual: A step by step guide to data analysis using SPSS for Windows (3rd ed.). Glasgow, UK: Bell & Brain. https://doi.org/10.1111/1753-6405.12166

Phipps, D., Meakin., G., & Beatty, P. (2011). Extending hierarchical task analysis to identify cognitive demands and information design requirements. *Applied Ergonomics, 42*(5), 741-748. https://doi.org/10.1016/j.apergo.2010.11.009

Plonsky, L., & Oswald, F. L. (2014). How big is 'big'? Interpreting effect sizes in L2 research. Language Learning, 64(4), 878–912. https://doi.org/10.1111/lang.12079

Polio, C., & Williams, J. (2011). Teaching and testing writing. In M.H. Long & C.J. Doughty (Eds.), *Handbook of language teaching* (pp. 486-517). Malden, MA: Blackwell. https://doi.org/10.1002/9781444315783.ch26

Pujola, J. (2001). Did CALL feedback feed back? Researching learners' use of feedback. ReCALL 13(1), 79-98. https://doi.org/10.1017/s0958344001000817

Ramaswamy, N. (2012). Online tutor for research writing. (Unpublished thesis). Retrieved from Iowa State University, ProQuest Dissertations Publishing, 1519246. https://doi.org/10.31274/etd-180810-2048

Rapp, C., & Kauf, P. (2018). Scaling academic writing instruction: Evaluation of a scaffolding tool (Thesis writer). *International Journal of Artificial Intelligence in Education, 28*(1), 1–16. https://doi.org/10.1007/s40593-017-0162-z

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2007). Designing for metacognition—applying cognitive tutor principles to the tutoring of help seeking. *Metacognition and Learning*, *2*(2-3), 125–140. https://doi.org/10.1007/s11409-007-9010-0

Roscoe, R. D., Snow, E. L., & McNamara, D. S. (2013, July). Feedback and revising in an intelligent tutoring system for writing strategies. In *International Conference on Artificial Intelligence in Education* (pp. 259-268). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-39112-5_27

Sándor, Á. (2007). Modeling metadiscourse conveying the author's rhetorical strategy in biomedical research abstracts. *Revue Française de Linguistique Appliquée*, *12*(2), 97–108. https://doi.org/10.3917/rfla.122.0097

Sawyer, R. K. (Ed.), *The Cambridge handbook of the learning sciences*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1192/bjp.bp.106.029678

Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language. A case study of an adult learner of Portuguese. In R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237-326). Rowley, MA: Newbury House. https://doi.org/10.2307/326454

Schwartz, D.L., & Bransford, J.D. (1998) A Time For Telling, Cognition and Instruction, 16:4, 475-5223. https://doi.org/10.1207/s1532690xci1604_4

Sharma, P. & Hannafin, M.J. (2007). Scaffolding in technology-enhanced learning environments. *Interactive Learning Environments, 15*(1), 27-46. https://doi.org/10.1080/10494820600996972

Strobl, C., Ailhaud, E., Benetos, K., Devitt, A., Kruse, O., Proske, A., & Rapp, C. (2019). Digital Support for Academic Writing: A Review of Technologies and Pedagogies. *Computers & Education, 131*, 33-48. https://doi.org/10.1016/j.compedu.2018.12.005

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.

Swales, J.M. (1981). Aspects of articles introductions. *Aston ESP Reports, No. 1.* The University of Aston in Birmingham.

Swales, J.M. (1990). *Genre analysis: English in academic and research settings.* Cambridge: Cambridge University Press. https://doi.org/10.1075/z.184.513swa

Swales, J.M. (2004) Research Genres: Exploration and Application. Cambridge University Press, Cambridge. http://dx.doi.org/10.1017/CBO9781139524827

Tardy, C. M. (2009). *Building genre knowledge*. West Lafayette, IN: Parlor Press.

Vandermeulen, N., Leijten, M., & Van Waes, L. (2020). Reporting writing process feedback in the classroom: Using keystroke logging data to reflect on writing processes. *Journal of Writing Research, 12*(1), 109-140. https://doi.org/10.17239/jowr-2020.12.01.05

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education, 16*(3), 227–265.

Vygotsky, L. S. (1962). *Thought and Language.* Cambridge MA: MIT Press. https://doi.org/10.1017/s0272263100008172

Appendix A1: Definitions of the moves and steps of research article Introductions

Create-A-Research-Space model based on Swales (1981, 1990, Author1 et al., 2015)

Moves	Steps	Definitions
Move 1: Establishing a Territory	The purpose of Move 1 is to foreground what is generally known and what has been previously researched by elaborating on the current state of the art on the topic, emphasizing what makes the topic central in the field, emphasizing the significance of the research agenda, and reviewing literature. The content of this move blends theoretical, empirical, and general, common-knowledge information.	
	Claiming centrality	Highlights the prominence, importance, and interest in the target topic exhibited by the research community in order to: ▪ substantiate the worthiness of investigating the topic ▪ entice the interest of the reader.
	Providing general background	Overviews the topic knowledge space and provides background to the study by presenting generally known information on the topic with or without referencing literature (e.g., non-empirical publications like government documents, guidelines, laws, general statistics, etc.) in order to: ▪ build an informational, conceptual, or theoretical frame of reference for the reader ▪ link to common knowledge related to the topic ▪ present an evaluative stance towards related common knowledge ▪ indicate a broader scope for the reported study ▪ support the reader's understanding of the reported study.
	Reviewing previous research	Reviews relevant literature by synthesizing previous research in order to: ▪ build an empirical and/or theoretical frame of reference for the reader ▪ show what shared empirical and/or theoretical knowledge builds the current state of the art on the topic

		▪ demonstrate knowledge and understanding of the topic ▪ acknowledge what has been found or claimed ▪ attribute empirical findings to their authors ▪ provide a stance towards previous findings ▪ support understanding and credibility of the study to be presented further.
Move 2: Identifying a niche	The purpose of Move 2 is to flag problems, limitations, incompleteness, and/or gaps in the research territory. It serves to justify the need for a new research contribution to the field.	
	Indicating a gap	Specifies the unknown in the targeted research agenda or domain of practice that needs to be filled in order to: ▪ show connections between what is known in the field and what requires investigation ▪ demonstrate critical evaluation of the current state of the art.
	Highlighting a problem	Specifies problems/challenges in the targeted research agenda or domain of practice that require attention and/or improvement in order to: ▪ signal existing issues ▪ raise concern about existing issues ▪ demonstrate critical evaluation of the current state of the art.
	Raising general questions	Raises general questions based on the knowledge territory and/or based on the identified gap or problem in order to: ▪ show authors' critical reflection ▪ bolster up interest in the topic ▪ possibly connect to general hypotheses and/or speculations ▪ possibly connect to the research questions of the current study.
	Proposing general hypotheses	Puts forth general hypotheses about possible future findings or implications based on the existing body of knowledge, and/or based on the knowledge territory, and/or based on the identified gap or problem in order to: ▪ provide a possible explanation to the unknown

		▪ introduce a tentative foundation for possible investigation ▪ imply that there is insufficient evidence to support topic-related claims ▪ possibly connect to the research questions or hypotheses of the current study.
	Presenting justification	Emphasizes and justifies the need to address the specified gap, problem, practical need, questions, and/or hypotheses that constitute the niche in order to: ▪ call for action ▪ substantiate the importance of conducting new research ▪ raise awareness of potential beneficial outcomes ▪ possibly transition to the reported study.
Move 3: Addressing the niche	The purpose of Move 3 is to show how the present study addresses the identified niche. It previews essential aspects of the study such as research goals, questions/hypotheses, methodology, and main results. Arguments about the value of the reported work and an outline of the content of the paper may be integrated in this move as well.	
	Introducing present research descriptively	Introduces the main features of the study by providing brief descriptive information in order to: ▪ acquaint the reader with the gist of the study ▪ situate the current study in the newly defined niche ▪ show continuity between the study and previous research ▪ possibly connect to the study's research goals/hypotheses.
	Announcing present research purposefully	Introduces the purpose of the study in order to: ▪ specify research goals ▪ set a clear stage for reporting the study in the following sections.
	Presenting research questions	Introduces the research questions of the study in order to: ▪ specify the points of inquiry ▪ connect the topic with specific points of inquiry ▪ set a clear stage for reporting the study in the following sections.

	Presenting research hypotheses	Introduces hypotheses about findings that might be obtained in the study in order to: ▪ specify the assumptions to be tested in the study ▪ clarify expectations for the findings and potential outcomes.
	Clarifying definitions	▪ Defines terms or concepts as they are used in the study in order to: ▪ explain the meaning of used terminology ▪ provide working definitions and/or clarify operationalized constructs ▪ show how the terms or concepts used in the paper coincide with or diverge from similar concepts in the field ▪ prevent readers' misinterpretation of used terminology.
	Summarizing methods	Provides a succinct summary of the research design, procedures, and/or data analysis in order to: ▪ illustrate how the information needed to accomplish the purpose of the study was obtained ▪ point out the most important aspects of the methodology ▪ begin establishing credibility.
	Announcing principal outcomes	Briefly announces the principal findings of the study in order to: ▪ specify which findings help occupy the niche ▪ highlight important take-home messages.
	Stating the value of present research	Articulate the value of the study in order to: ▪ emphasize that the study offers a significant contribution to the field ▪ highlight important implications of the study.
	Outlining the structure of the paper	Previews the structure of the paper and/or content in order to: ▪ inform the reader of how the paper is organized ▪ guide the reader through the content of paper ▪ allow the reader to easily locate specific content.

Appendix A2. Examples of the moves and steps of research article Introductions from the corpus in the Research Writing Tutor

Move 1: Establishing a Territory

Claiming centrality

"There is a considerable interest within the food and pharmaceutical industries in the creation of delivery systems that can encapsulate, protect, and deliver ω-3 fatty acids because of their potential health benefits in decreasing the risk of coronary heart disease, immune response disorders, ulcerative colitis and Crohn's disease, and mental illnesses (1-6)."

Providing general background

"These syndromes are characterized by diffuse alveolar damage."

Reviewing previous research

"Several lines of evidence suggest that phosphorylation of SLP-76 is critical to bring VAV1 and Nck into proximity (Raab et al, 1997; Bubeck Wardenburg et al, 1998; Zeng et al, 2003)."

Move 2: Identifying a niche

Indicating a gap

"An exhaustive survey of the literature reveals that no work has been reported on the reactions of different nitriles with 16-hydroxyhexadec-cis-9-enoic acid in strong sulfuric acid medium."

Highlighting a problem

"The one limitation to these findings regarding narrow-row soybean is that all the studies were done with non-glyphosate-resistant cultivars."

Raising general questions

"Still, can it be claimed that this method can reduce spherical aberration to acceptable levels?"

Proposing general hypotheses

"The extent to which a message is perceived as having an impact on the audience is likely to accentuate the perceived harm to the victim."

Presenting justification

"To build a useful understanding of a social network, a complete and rigorous description of a pattern of social relationships is a necessary starting point for analysis."

Move 3: Addressing the niche

Announcing present research descriptively

"In this work, we characterize the dynamics of the backbone and side chain segments of the colicin Ia channel domain on the time scales of picoseconds to milliseconds and compare the soluble and membrane-bound states."

Announcing present research purposefully

"The aims of this paper were to quantify the effects of social stressors on the performance of growing pigs, including variation in their ability to cope, and to incorporate these relationships into a more general growth model (Wellock et al., 2003a) to allow the prediction of more complex interactions."

Presenting research questions

"To explore any differences between the participation protocols, two sub-questions guided our study: (1) Did social construction of knowledge occur in both forums?; (2) Did the participation protocols affect knowledge construction and participation?"

Presenting research hypotheses

"We predict that, in contrast to these possible effects of attitude training on motivation, positive attitudes and the motivational processes that they may cue will not supply the cognitive capacity one needs to excel on complex cognitive tasks completed under stereotype threat."

Clarifying definitions

"Illegal logging is defined here as all the activities related to the selective harvesting and transport of logs from natural forests that are made in contravention to national regulations."

Summarizing methods

"Experiments were performed using the bagasse, as it comes from an alcohol/sugar factory, and bagasse screened size 0.248 to 1.397 mm (12–60 mesh) to evaluate the possibility of using the bagasse as it comes from the mills."

Announcing principal outcomes

"The results show that, contrary to previous suggestions, m-calpain is not extensively degraded during postmortem storage but becomes increasingly bound to the myofibrillar fraction and that m-calpain isolated from 7-d postmortem muscle is proteolytically inactive."

Stating the value of present research

"This paper extends and deepens this growing international accounting literature by reporting detailed evidence of the links between earnings discontinuities and accruals manipulation based on a large sample of UK firms."

Outlining the structure of the paper

"The remainder of the paper is organized as follows. Section 2 describes the LIS software and the interface to WRF. Section 3 presents the experiment design for

comparing LIS-initialized WRF runs to control simulations. Results and composite verification statistics are presented in section 4. Finally, section 5 summarizes the results and discusses possible follow-up research."

Appendix B. Distribution of steps in the Introductions of published research articles in the Research Writing Tutor corpus

The first pie chart represents the distribution of steps in all 900 Introductions in 30 disciplines.
The other pie charts represent the disciplines of the students who participated in this study.