2015

Automated Writing Analysis for writing pedagogy: From healthy tension to tangible prospects

Elena Cotos

Iowa State University, ecotos@iastate.edu

Follow this and additional works at: http://lib.dr.iastate.edu/engl_pubs

Part of the Curriculum and Instruction Commons, Educational Assessment, Evaluation, and Research Commons, Educational Methods Commons, and the Higher Education Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/engl_pubs/63. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.
Automated Writing Analysis for writing pedagogy: From healthy tension to tangible prospects

Abstract
This article aims to engage specialists in writing pedagogy, assessment, genre study, and educational technologies in a constructive dialog and joint exploration of automated writing analysis as a potent instantiation of computer-enhanced assessment for learning. It recounts the values of writing pedagogy and, from this perspective, examines legitimate concerns with automated writing analysis. Emphasis is placed on the need to substantiate the construct-driven debate with systematic empirical evidence that would corroborate or refute interpretations, uses, and consequences of automated scoring and feedback tools intended for specific contexts. Such evidence can be obtained by adopting a validity argument framework. To demonstrate an application of this framework, the article presents a novel genre-based approach to automated analysis configured to support research writing and provides examples of validity evidence for using it with novice scholarly writers.

Disciplines
Curriculum and Instruction | Educational Assessment, Evaluation, and Research | Educational Methods | Higher Education

Comments
This is a manuscript of an article from Writing and Pedagogy 6 (2015). Posted with permission.
Abstract

This article aims to engage specialists in writing pedagogy, assessment, genre study, and educational technologies in a constructive dialog and joint exploration of automated writing analysis as a potent instantiation of computer-enhanced assessment for learning. It recounts the values of writing pedagogy and, from this perspective, examines legitimate concerns with automated writing analysis. Emphasis is placed on the need to substantiate the construct-driven debate with systematic empirical evidence that would corroborate or refute interpretations, uses, and consequences of automated scoring and feedback tools intended for specific contexts. Such evidence can be obtained by adopting a validity argument framework. To demonstrate an application of this framework, the article presents a novel genre-based approach to automated analysis configured to support research writing and provides examples of validity evidence for using it with novice scholarly writers.

“Like any other technology, writing assessment and the techniques and artifacts associated with it fall prey to missing the forest for the trees when it comes time to evaluate how such technologies are put into use.” (Elliot Knowles)

Introduction

In the contemporary world, educational and workforce establishments require evidence of strong written communication skills, expecting such evidence to be efficiently generated for timely decisions. This global demand has put a strain on the writing teacher community. Their mission to help students develop as writers has never been an easy one, especially given the time-consuming grading and copious commenting on hundreds of student papers that writing teachers are required to do. To this end, it seems daunting for a teacher with such a load to individualize instruction to the extent where every student would be provided with sufficient writing opportunities, each accompanied by teachers’ evaluation and guidance for the multitude of aspects needing follow-up practice and improvement.

Attempting to relieve practitioners of this burden, assessment-driven developers of education technologies have been promoting automated writing analysis (AWA) systems, rationalizing their helpfulness and suitability with statistical evidence of human-computer agreement. In response, writing teachers and researchers have articulated severe criticism against AWA, even requesting to expel it from the writing classroom (Herrington & Moran, 2012). Thus, instead of being received as “but one tool to help students and their instructors along the way,” (Klobucar,
Deane, Elliot, Ramineni, Deess, & Rudniy, 2012: 105), this technology seems to have created a widening gap rather than a link between pedagogy and assessment.

This article discusses the nature of the problem, first examining the pedagogical values and the kernels of discontent with AWA, which has been discredited for its inability to read for meaning and for measuring the writing construct in a very restricted way. The intent here is not to enter the debate created by the lack of alignment between construct representation views and controversial computational realizations of aspects of writing. Rather, I would like to acknowledge legitimate concerns and to emphasize the need to substantiate the construct-driven discussion with appropriate empirical evidence of how AWA may reinforce learning and assessment. If framed as a validity argument (Kane, 2006), such evidence can be gathered consistently to corroborate or refute interpretations, uses, and consequences intended for a given context. Most importantly, adopting a systematic validity framework, especially when automated analysis is intended to complement teaching, would allow stakeholders to engage in collaborative efforts for AWA evaluation and improved design to better support learning to write. As an example, I will describe how automated analysis can be configured for a specific instructional purpose, demonstrating a direction for future research and development through a genre-based approach.

Fundamentals of Automated Writing Analysis

Technological innovations have steadily entered education, exerting profound effects on teaching and learning. Many educational technologies face barriers in their adoption, mandating that developers, researchers, and practitioners establish a clear link between the use of technology and the different ideologies of teaching and learning, which govern the expectations of usefulness and effectiveness. AWA is no exception in that it has powered paradigmatic shifts in writing assessment and instruction, and raised both enthusiasm and reasonable concerns.

**Automated scoring and evaluation**

Judgment of the prospects and restraints of AWA requires a sense of the field’s axiology. AWA stems from empirical methods for linguistic description and modelling, its evolution dating back to the 1960s when the Project Essay Grade (PEG) system was pioneered (Page, 1994). Since then, AWA offspring have been continuously invented to generate scores and feedback similar to human ratings. In general, AWA systems rely on scoring algorithms and are grounded in cognitive information-processing. Different natural language processing and statistical methods are exploited to identify linguistic features reflecting various aspects of texts that are relevant to the writing construct. Leaders in AWA development, Burstein, Tetreault, and Madnani (2013) explain that these automatically-generated linguistic features are viewed as “tangible markers in essay writing that can be used to measure (evaluate) writing quality, given a specific writing task” (p. 57). AWA, thus, operationalizes writing as containing features, which, when measured quantitatively, can help make relatively accurate predictions of holistic human evaluation, considering “organization and development of ideas, the variety of syntactic constructions, the use of appropriate vocabulary, and the technical correctness of the writing in terms of its...
grammar, usage, and mechanics” (p. 57). These features derive from scoring rubric criteria specific to writing tasks.

Before I canvas the AWA controversies, an important distinction needs to be made within AWA between Automated Essay Scoring (AES) and Automated Writing Evaluation (AWE). This distinction corresponds to the summative and formative uses of writing assessment, the two positing differences in function and timing. Summative assessment measures students’ performance at the end of a course or a program with the purpose of grading, certification, or accountability. Formative assessment is conducted during the teaching process for the purpose of identifying students’ difficulties, scaffolding learning, and monitoring the appropriateness of instruction (Brown & Knight, 1994; Ebel & Frisbie, 1991; Torrance & Pryor, 1998).

AES for assessment and AWE for instruction
AES is represented by automated scoring engines like e-rater, Intelligent Essay Assessor (IEA), Intellimetric, Constructed Response Automated Scoring Engine (CRASE), AutoScore, Bookette, Mosaic, Accuplacer, etc., which were designed for immediate, individualized, and consistent scoring to complement summative assessment by human raters (e.g., in large-scale tests including the Internet-based Test of English as a Foreign Language - TOEFL iBT®, Pearson Test of English - PTE, and Graduate Management Admissions Test - GMAT). In educational measurement, perhaps the most central concern for any type of assessment is validity, defined as “an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment” (Messick, 1989, p. 13). Given that the scoring engines are modelled on human ratings guided by scoring rubric criteria, a notable focus in AES validation research has been on criterion validity in terms of agreement between automated and human scores. Reported correlations and agreement rates are relatively high; cross-validated correlations between the scoring engines generally range from .70 to .90, often being in the .80–.85 range (Dikli, 2006; Elliot, 2013; Landauer, Laham, & Foltz, 2003; Page & Petersen, 1995; Shermis & Hammer, 2013). Automated scores appear to behave well when compared with other standardized tests and some non-test measures (Keith, 2003; Weigle, 2010, 2013a). Construct validity and its association with instructional activities have also fallen within the scope of AES system-centric research (Attali & Burstein, 2006; Page, Keith, & Lavoie, 1995). In the past decade, an important trend examining AES within the larger context of the argument-based approach to test validation (Kane, 1992, 2006) has focused on applying, refining, and expanding conceptual validation frameworks to particular applications of automated scoring (e.g., Bennett & Bejar, 1998; Xi, 2008; Williamson, Xi, & Breyer, 2012).

AWE systems developed for formative assessment such as Criterion, WriteToLearn, MyAccess!, Writing Power, Writing Roadmap, etc., have shown a parallel stream of evolution. These are complex applications of the scoring engines for the generation of automated feedback, which typically addresses grammar, syntactic complexity, mechanics, style, topical content, content development, deviance, etc. Like their AES siblings, AWE programs have seen large-scale implementations in writing classrooms in elementary, middle, and high schools; community colleges; universities; job training programs; and military institutions (Burstein, Chodorow, & Leacock, 2004). Following calls to validate contextualized uses of AWE by students and teachers in order to discern the intricacies of classroom ecologies and to gain an understanding of how AWE can facilitate learning and instruction (Warschauer & Ware, 2006), the AWE research agenda expanded to naturalistic classroom-based investigations of how, when,
if, and why this technology works. Consequently, evidence of effectiveness began to be sought from users’ experiences and outcomes. User-centric results show varied student/teacher perceptions, increased student motivation and learner autonomy, positive feedback effects, as well as enhanced understanding of errors and writing improvement (Attali, 2004; Chen & Cheng, 2008; Chung & Baker, 2003; El Ebyary & Windeatt, 2010; Elliot & Mikulas, 2004; Foltz, Laham, & Landauer, 1999; Grimes & Warschauer, 2008; Li, Link, Ma, Yang, & Hegelheimer, 2014; Link, Dursun, Karakaya, & Hegelheimer, 2014; Rich, Schneider, & D’Brot, 2013; Schroeder, Grohe, & Pogue, 2008; Ware, 2011, 2014). Although the increasing number of studies sheds positive light on the potential of AWE, evidence accumulated to support its validity is relatively inconsistent and, hence, insufficiently informative for use in writing classrooms.

### Writing praxis

First and second language (L1 and L2) writing courses are generally housed by departments and programs in English, Linguistics, and Education, writing instruction often being provided by specialists in Rhetoric and Composition and Applied Linguistics. Although rooted in distinct philosophical beliefs, these fields share a common concern with pragmatic applications and are anything but unallied and disconnected. Intersecting rather notably, particularly via contributions of scholars whose work crosses disciplinary boundaries (e.g., J.M. Swales and P.K. Matsuda), they experienced similar transformational changes over the years and have come to develop mutual trajectories and overarching tenets governing writing praxis.

#### Principles guiding classroom praxis

As Hyland (2003: 1) puts it, “[e]verything we do in the classroom, the methods and materials we adopt, the teaching styles we assume, the tasks we design, are guided by both practical and theoretical knowledge, and our decisions can be more effective if that knowledge is explicit.” While reviewing the theoretical landscape that undergirds modern day L1 and L2 writing praxis is beyond the scope of this article, it is essential to consider a summative representation of theoretically-grounded practices. The following recommendations are a rather tall order synthesis, for their implementation requires well-resourced and highly-trained teachers; yet, they help make explicit the fundamental principles proposed to enact best practices. In short, students need to be provided with writing instruction that systematically:

- exercises purposeful social/cultural/disciplinary activities for writing as learning and discovery;
- develops rhetorical knowledge by learning and practicing key concepts such as audience, purpose, context, and genre;
- integrates exposure to and analysis of different genres and conventions;
- draws on analyses of authentic discourse;
- fosters conscious control of language use to achieve communicative purposes and intended effects on the audience;
- provides context-specific formative feedback frequently and timely;
- allows time for recursive practice and iterative revision;
- creates conditions for cognitive and metacognitive operations by implementing flexible process writing strategies;
- stimulates interpersonal and intrapersonal interaction;
exploits the mediating role of various types of scaffolding resources;
reflects discipline-specific expectations and prepare for writing in the discipline;
affords writing in various environments, including collaborative and digital;
provides opportunities for developing writing as a language skill.

These principles underscore the value of cultivating an understanding of target contexts, audiences and expectations, and highlight the importance of engaging in a process of productive practice leading to communication that is both rhetorically and linguistically effective. These ideas resonate with the construct of writing construed by genre pedagogy, where writing is viewed as a culturally and socially mediated behaviour.

**Writing pedagogy and genre**

A longstanding movement has contributed to theorizing the role of genre in the creation and interpretation of texts and cultures. Scholarship produced by researchers in Rhetoric and Composition, Applied Linguistics, Education, and Communication has helped transform genre study from a descriptive to an explanatory activity, one that investigates not only text-types and classification systems, but also the linguistic, sociological, and psychological assumptions underlying and shaping these text-types (Bawarshi, 2000: 335). Having established itself as theoretically robust, linguistically informed, and evidence-based, genre study has had a considerable uptake in writing classrooms (Hyland, 2007), where teachers are increasingly drawing on texts that students are expected to produce in academic, professional, and social contexts.

In Rhetoric and Composition, genre is defined as “typified rhetorical actions based in recurrent situations” that over time have been conventionalized depending on the needs, values, and ideologies shared by the target discourse communities (Miller, 1984, p. 159). Therefore, learning to write for an academic, professional, or workplace audience implies mastering its rhetorical conventions, and becoming a legitimate member of a discourse community involves engaging in social action adhering to its preferred communicative practices. From this perspective, genre-based pedagogy, particularly for L1 writing, embraces classroom inquiry and critical reflection about the values perpetuated by genres, social contexts, ideological climates, audiences with respective communication constrains, writing purposes, implied power relations, and different conditions of genre use that may influence rhetorical structures and writers’ choices – all leading to student critique and negotiation of genres for the purpose of aligning them with their immediate needs and purposes (Johns, 2011; Kostouli, 2009). Exposure to genres is intended to provide learners with opportunities for understanding the multiple voices that populate genres (Bakhtin, 1981) in order to experience the multiple facets of the social encounter inherent in genre practice, which can be shaded by “power, authority, meaning making, and identity that are implicit in the use of literacy practices within specific institutional settings” (Lea & Street, 2006: 370).

Within Applied Linguistics, two genre traditions have been most influential on L2 writing pedagogy: Systemic Functional Linguistics (SFL) and English for Specific Purposes (ESP). Both of these bourgeoning fields accentuate the role of language and make an explicit connection between texts and contexts. SFL treats language as a social semiotic central to communicative activity (Halliday, 1978). It focuses on texts and textual features, accounting for the purposeful and successive character of genres, the textual patterns underlying socially recognized functions, and the relation between language and context (Halliday, 1985). Concerned with how language
serves communication, SFL-based pedagogy foregrounds the role of functional language, or “language that is doing some job in some context” (Halliday & Hasan, 1989: 10). Contexts are viewed as interactive events of social exchange of meanings 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: 11). In other words, contexts define how meaning potential is realized in texts, determining the language choices made to express particular meanings.

The analogous ESP genre-oriented approach is a fusion of research and pedagogy, which generates foundational knowledge about communication in specific contexts to facilitate the teaching and learning of context-specific language that non-native speakers need to acquire in order to successfully engage in targeted social practices. The broader field of ESP gave rise to English for Academic Purposes (EAP), which has had a momentous impact on writing instruction. Drawing from a rich research base and an assortment of theoretically grounded techniques, EAP is concerned with the communicative needs and practices of English language learners in academic contexts in view of the cognitive, social, and linguistic demands of their disciplines (Hyland & Hamp-Lyons, 2002: 2). Via ESP/EAP, L2 writing instruction has witnessed an immense imprint of genre theory. Merging linguistic phenomena and social context dynamics, genre-based pedagogy exposes students to various intricacies of the social dimension by revealing to them different ways in which the conventions that build texts “produce the ‘whys’ of social effect” (Cope & Kalantzis, 1993: 8). It provides apprenticeship into the prototypical discourse patterns of literate practices associated with genres, “offering learners substantial practice in analyzing linguistic and rhetorical norms that typify a Discourse’s common text forms while promoting the cognitive skills needed to decode and reproduce these text forms” (Ferris & Hedgcock, 2014: 72).

EAP practitioners, especially those working with advanced L2 graduate writers, have widely adopted Swales’ (1981, 1990, 2011) genre analysis framework of ‘moves’ and ‘steps.’ The moves are “bounded communicative act[s] […] designed to achieve one main communicative objective” (Swales & Feak, 2000: 35), and the steps are rhetorical strategies that convey specific communicative functions. This EAP genre approach is particularly advocated as “a more situated pedagogy for novice writers” (Polio & Williams, 2011: 498), who are recommended “to begin with text structures and then to move rapidly to viewing genres as socially mediated entities” (Johns, 2011, p. 64). L2 writing teachers have also welcomed corpora, or principled collections of texts, as resources for cultivating learners’ genre awareness and acculturation into target discourses.

In addition to the social-constructionist genre views, L2 writing instruction is steered by second language acquisition (SLA) theory. The SLA Interaction Approach, in particular, emphasizes the role of learners’ internal processes (Long, 1996) and is thus congruent with L1 cognitive models of writing development (Flower & Hayes, 1981). Describing the processes invoked when the learners encounter input, engage in interaction, receive feedback, and produce output, the Interaction Approach explains how interaction and learning in a linguistic environment can be linked to the cognitive concepts of noticing, working memory, and attention (Gass & Mackey, 2006, p. 176). Input, or the target language to which learners are exposed, has to draw their attention to what is linguistically accurate and culturally appropriate, as well as to what is ungrammatical or unacceptable (Gor & Long, 2009). To make such input information available to students and induce them into noticing the peculiarities of linguistic phenomena, teachers assign tasks that create opportunities for inter- and intra-personal interaction.
Objections to AWE

AWA developers, including commercial vendors and independent researchers, genuinely aim to facilitate the jobs of various stakeholders with what they believe are efficient and effective computational solutions. The time and cost-effectiveness of automated analysis generally appeal to educational institutions, which, often under the pressure of common standards, tend to welcome immediate and reliable computer-based scoring as a viable supplement to human evaluation. AWE programs are promoted as effective enhancers of process writing instruction, compelling guides for student revision, and robust vehicles of consistent writing and evaluation across the curriculum. They are also presumed to motivate multiple drafting and revision, foster learner autonomy, and enhance the instructional dynamic by supporting the drive toward individualized instruction (Attali & Powers, 2008; Burstein, 2012).

Unfavourable prospects

Despite these purported benefits, writing teachers express reluctance to implement AWE in the classroom, primarily because they feel “Writing to a machine is not writing at all” (Herrington & Moran, 2012: 219). A number of issues have been raised, casting doubts on the place of these technologies in the future of writing instruction (Byrne, Tang, Truduc, & Tang, 2010; Elliott, 2011; Ericsson & Haswell, 2006; Herrington & Stanley, 2012; Jones, 2006; McCurry, 2010; Perelman, 2012a). Following is a synthesis of views expressed in the literature as well as the professional voice of the writing community expressed in Position Statements of the Conference of College Composition and Communication (CCCC, 2004, 2009) and of its parent, the National Council of Teachers of English (NCTE, 2013):

- scoring systems cannot read for meaning, identify communicative intent, evaluate the argumentation quality, and verify factual correctness;
- machine analyzers are calibrated to static features and formulaic expression, thus heavily subordinating meaning;
- scoring systems decontextualize writing, depriving it of the social and communicative dimensions and eliminating the value of human audiences in real-world contexts;
- computers cannot replicate the cognitive processes activated in the brain when humans make evaluative judgments about the quality of writing;
- students may consciously or unconsciously adjust their writing to match the assessment criteria of the software;
- teachers may feel pressured to support such adjustment in an attempt to raise test scores;
- students may trick the software, and thus the assessment would capture their ability to use machine-tricking strategies rather than writing ability;
- scoring systems designed for L1 writers may leave developing L2 writers at a disadvantage;
- pre-programmed essay prompts may de-skill teachers by inhibiting them from devising creative writing and assessment opportunities for their students;
- different types of bias may be veiled under the ‘black-box’ models, the scoring criteria of which are unclear to lay audiences.

Clearly transgressing the principles that the writing community has come to embrace as well as the expectations inherent to genre-based pedagogies, these issues are pointed out again and again
in discussions within professional organizations, among individual scholars, and in the popular press. Overall, fundamental objections are directed at measuring the writing construct with no regard to writing as social and communicative practice. This major criticism is enmeshed with issues arising from the lack of vital intersections between teaching, assessment, and technology.

**Origins and ramifications**

**Ontogeny**

On its path to advancement, writing pedagogy has acquired a non-product stance. However, AWE computational operationalization with a focus on such features as grammar, usage, and mechanics may seem to revive the product-oriented approach. The idea that written performance can be evaluated based on predictive models may be reminiscent of the cognitivists’ attempts to predict and generalize the writing process. Additionally, the importance of genre appears to be reduced to essay text types only, as if this is the only genre developing writers need to master.

Most of the concerns regarding AWE share a unified opposition to automated scoring. Writing instruction and assessment have hardly ever seen eye to eye, assessment being a contested space, and “[n]owhere is this tension more apparent than in the contrasting ways to stump automated assessment writing systems” (Elliot & Klobucar, 2013: 16). Neal (2011: 68) cautions: “When companies bolster their use of machine scoring with claims of high levels of correlation with human readers, we should immediately see warning flags and remember the similar way they continue to promote indirect measures of writing in the face of all the arguments to contrary.” Writing practitioners see the evaluation offered by AES as contradicting their views and values. To them, assessment is a means of improving teaching and learning. It should be tailored to the primary purpose of use, be context-specific and needs-based, elicit contextualized and meaningful writing, and capture a variety of integrated skills (CCCC, 2009; NCTE, 2013).

**Construct**

A preeminent conundrum extending from the assessment angle is the representation of the writing construct, i.e. “the way writing is understood by a given community (Elliot, Ruggles Gere, Gibson, Toth, Whithaus, & Presswood, 2013). Educational measurement research informing the design of AWA, relying on a correlationally-grounded notion of concurrent validity, view writing as a construct containing quantifiable features that, “in the aggregate, embody the meaning of writing for the assessment” (Williamson, 2013: 166). The writing community objects to such metrics, arguing that the writing construct is essentially about meaning making to achieve a specific communicative purpose and, thus, cannot be broken down into formulas and evaluated computationally. From their perspective, the writing construct includes “the rhetorical ability to integrate an understanding of audience, context, and purpose when both writing and reading texts; the ability to think and obtain information critically; the ability to effectively employ multiple writing strategies; the ability to learn and use the conventions appropriate to a specific genre of writing; and the ability to write in various and evolving media” (Perelman, 2012b: 129). It must be noted that these descriptors of the writing construct link construct representation to both cognitive and interpersonal dimensions in any given instance of assessment (Elliot et al., 2013), which current state-of-the-art AWA seems not to account for.

**Washback**
Assessment (be it summative or formative) is never context-free, so many of the concerns mentioned above are related to washback, or “the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit […] learning” (Messick, 1996: 241). In the dispute between AWE supporters and opponents, little consideration is given to contextual factors that may cause one or another kind of washback. Existing research has tangentially documented that AWE programs are often misused (Attali, 2004; Shermis & Burstein, 2013; Warschauer & Grimes, 2008), implementation choices varying considerably in scope and approach. It is not uncommon for teachers to adopt AWE for scoring or test preparation purposes, reserving very little time for revision and neglecting these systems’ formative feedback capabilities. Some teachers, however, employ AWE to complement more germane types of activities such as pre-writing, writing practice, peer review, revision, and teacher commenting. The outcomes in these cases can be more satisfactory both for students and teachers (Li, Link, & Hegelheimer, 2015; Warschauer & Grimes, 2008). Although teachers do not seem to dislike AWE, they may rarely use it in their classrooms, especially because some writing tasks are representative of genres outside the software’s analytical capabilities.

Validity
The feedback of AWE programs is a direct output of scoring systems; in other words, the linguistic features used for scoring are disaggregated and applied to feedback. The very assumption that such direct transfer from automated scoring to formative assessment is suitable has raised validity concerns (Cotos, 2012; Keith, 2003). Moreover, claims of AWE validity have almost exclusively been based on psychometrically-driven evidence of AES reliability (Warschauer & Ware, 2006). While reliability studies supply evidence in support of automated scoring for summative assessment, they do not satisfy the validity requirement for assessment for learning where much regard must be given to the ecologies of various target contexts. Considering these justifiable criticisms, AWE research needs to take a direction opposite to the de-contextualized focus of AES research. As validity evidence “can include but is not limited to various forms of reliabilities” (Bachman, 1990: 96), conclusions about AWE effectiveness need to be derived from multifaceted, relevant empirical backing integrated in a manifold validity argument.

Non-inclusion
Another problem is that writing specialists were not involved in the conceptualization and development of the AWA technologies, which were introduced to them mainly as products that can assist with cumbersome writing assessment tasks. This approach is reminiscent of the nature of theory-to-practice relation in writing scholarship, with practice being at the receiving end (Zhu, 2010: 214). In the relationship between AWA and writing praxis, pedagogy still remains an end rather than an informer for technology design and application. Connections need to be made at multiple levels; Figure 1 indicates the missing links with dashed lines.
**Need for continued dialogue**

Many researchers have argued that the debate over the pros and the cons of AWA is often counterproductive. Whithaus (2006: 167) remarks that the “discourse of rejection” does not offer a tenable path. It is not the technology that should be condemned but rather “the vision of writing and assessment [...] in an era of testing and accountability” (Vojak, Kline, Cope, McCarthey, & Kalantzis, 201: 108) and “the practices that have led to the development and use of AES as we know it today” (Condon, 2013, p. 105). Similarly, Deane (2013: 12) suggests that “objections to AES are actually objections to the assumptions of standardized testing.” He argues that taking polar positions is a mistake; all involved should realize that “AES is one instantiation of a larger universe of methods for automatic writing analysis” and that it is important to consider “how forms of automated writing evaluation fit into education and assessment” (Deane, 2013: 8).

Artificial intelligence models and automated analysis algorithms will likely never be capable of truly understanding texts. However, they can be devised to assist in the construction of substantive and procedural knowledge for writing (Scardamalia & Bereiter, 1987) by prompting, guiding, and providing just-in-time assistance, which could be focused on increasing cognition or writing skills – depending upon specific learning and instructional goals. The creators of AWA thus need to invite L1/L2 writing scholars and teachers to contribute to the conceptual design of both summative and formative systems for written literacy. AWE could undergo tremendous improvement with expert input containing rationales as to why and how particular affordances should be created rather than what features are amenable to computational modelling. At the same time, if practitioners are engaged in collaboration, they will better understand the criteria used for computational operationalization, as well as its technical advantages and disadvantages. This will enable them to set rational and pragmatic expectations and to develop strategies that would compensate for existing limitations. Teachers could perhaps propose ways to enhance the mediating role of AWE by developing scenarios of interaction with the feedback and demonstrating how students could productively react to it at a given stage in the writing process.
As Ramineni and Williamson (2013: 37) see it, a “healthy tension” will continue, but if it is constructive, it can lead to expanding automated analysis approaches and redefining the writing construct currently used in computational modelling to account for writing as a complex activity “where reasoning skills, writing processes, genre practices, and the cultural and social contexts in which genres develop take center stage” (Deane, 2013: 9). Therefore, the AWA dispute needs to be revisited with an open-mind. It is imperative that both parties work together to ensure that AWA applications: a) reinforce good teaching and assessment practices; b) develop a transparent definition and a mutual interpretation of the writing construct that would bridge the two epistemologies to inform proper algorithmic operationalization; c) obtain validity evidence indicating how this technology could realize not only its technical potential but also its potential to aid learning; and d) develop user training approaches to ensure appropriate implementation and avoid negative washback.

Outlook for the future

Systematic empirical evidence is needed to provide rich insights for future AWA development and application, which should ultimately conciliate writing instruction and assessment to the benefit of developing writers.

Need for empirical evidence

Many arguments against the use of automated analysis of writing are sound and reasonable, but they have yet to be adequately rebutted essentially because the findings obtained from one-time trial submissions can hardly be generalized and accepted as convincing validity evidence. A frequently cited example is a review of Criterion, MyAccess! and WriteToLearn, which focused on the strengths and weaknesses of these AWE systems to promote writing as a socially situated activity (Vojak et al., 2011). This review relies on observations from reviewers trying out these programs or making judgements based on their demos. Vojak et al.’s (2011: 103) conclusions lack strength of evidence given their methodology, which is described as “our research team submitted a short essay in response to a prompt […]. We received a score of 4.3 out of 6. We then tacked onto our initial essay several unrelated paragraphs and received a score of 5.4 out of 6. […] In another instance, our research team wrote and submitted a nonsense essay […]”. In a similar fashion, Herrington and Moran (2012), who found that Criterion may point to errors that are not in fact errors, based their conclusions on the submission of one essay. The limitation of such reports is at least three-fold: their focus is confined to scoring accuracy; it is unclear whether the submitted essays were representative of target students’ writing; and this type of ‘testing to trick’ is not a writing task, nor a pedagogical approach that teachers would likely adopt in the classroom. Nevertheless, such work is eye-opening for AWE developers, who need to better understand the writing construct as interpreted by those who teach and research writing.

The process of validation should include evidence for justified use as well as the possibility of rebuttal. It would be more conclusive if approached through systematic evidence relevant to the interpretations, uses, and consequences of AWE, compiled under an argument-driven validity framework (Elliot, 2013; Xi, 2010). For example, Elliot et al. (2013) propose a conceptual model for validation of AWE use in local settings, intended as a heuristic for writing program administrators. Their model broadly encompasses the writing domain, which can include any genre and any possible setting, the writing construct adopted in a specific instructional site, and the writing construct as defined within a given AWE software. Elliot et al. (2013) then build on
this analytic model to review AWE research, showing the value of existing works in terms of theory, scoring, use, diversity, and consequence.

As a starting point for arguing about validity, it is necessary to formulate an interpretive argument comprising claims about intended purposes (see Chapelle, Enright, & Jamieson, 2008). For example, Chapelle, Cotos, and Lee (2015) formulate an interpretive argument for Criterion. Their paradigm consists of a chain of inferences about the interpretations and uses of automated feedback, propositional warrants associated with the inferences, and specific assumptions underlying respective warrants. Each inference is established when the warrant associated with it is sustained by a collection of backing evidence supporting the assumptions that underlie the warrant. From this perspective, Chapelle et al. (2015) provide evidence supporting the assumption that L2 students use automated feedback to make decisions on how to revise their drafts and correct errors. They also suggest a rebuttal, stating that students may not make sufficient or effective use of feedback because they may lack confidence in its accuracy (a theme also mentioned by Li et al., 2015). More broadly, interpretive arguments for AWE systems could include formulating and testing both claims based on principles guiding classroom praxis and potential rebuttals in terms of anticipated negative effects. More generally, this approach would ensure congruency in the validation and evaluation of AWA, needed to strengthen the connection with classroom practice. In what follows, I will exemplify select learner-centered data to illustrate evidence that can be garnered to build a validity argument for the utilization of a genre-based configuration of AWE.

**Genre-based Research Writing Tutor**

Developers have recognized that the “[e]valuation provided by AWE systems must be aligned with writing genres” (Burstein et al., 2014: 1). It has also been affirmed that this technology “can be deployed in innovative ways that might provide better support for writer cognition and integrate more fruitfully with the social practices that encourage quality writing” (Deane, 2013: 20). The Research Writing Tutor (RWT) is a new AWE tool that aligns with the purposes of a genre-based writing curriculum and thus adds a dimension that is responsive to a richer understanding of the writing construct relevant to genre-based pedagogy. Teaching objectives and learning needs, rather than computational modelling, shape the specifications of this program. It was designed for a graduate-level L2 writing course at a research-intensive university in the US, which is offered to students in various disciplinary programs. The course uses EAP methodology with a focus on Swalesian move/step conventions specific to the research article genre, aiming to help students develop academically compelling texts as expected by their disciplinary discourse community. While RWT does not necessarily resolve all the ecological challenges, it incorporates the characteristics that need to be accounted for in genre-based graduate writing instruction (see Cortes, 2007). Rather than giving a scoring-based “verdict,” the tool aims to facilitate students’ engagement in socio-disciplinary practice, drawing on the disciplinary conventions and representational resources of the research article genre. This type of engagement seamlessly blends in the understanding of the rhetorical problem, the awareness of conventions of knowledge construction established by the disciplinary community, and the ability to effectively convey content by connecting rhetorical purposes with the textual and linguistic features of the genre.

As summarized in Figure 2, RWT contains three Modules that integrate elements responsive to essential principles of writing pedagogy and conceptualized considering EAP, SFL, and SLA theoretical tenets. The “Understand Writing Goals” Module offers multimodal scaffolding of the move/step rhetorical concepts. “Explore Published Writing” exposes students to rhetorically
annotated corpora in their disciplines (Appendix A1), a concordancer that can be queried by steps within a given communicative move (Appendix A2), and corpora of original published manuscripts. “Analyze My Writing” provides a platform for students to write and iteratively submit their drafts for automated analysis in order to receive different forms of move and step-level feedback encouraging revision and improvement. All three modules exploit textual analysis to facilitate learning from disciplinary corpus data and to enhance the visibility of rhetorical and linguistic characteristics of the discourse by fostering students’ analytic reading of genre.

RWT enables non-linear navigation within and between the three Modules so that students can flexibly use them either separately or concurrently, depending on the stage in their writing process. Generally, teachers begin imparting students with genre knowledge using the “Understand Writing Goals” Module. Then, they assign corpus exploration activities with the “Explore Published Writing” Module to help students consolidate their declarative knowledge of genre conventions by observing how authors in their discipline compose discourse in the rhetorical ways explained in class. With this Module, RWT aids establishing an indirect but realistic connection to the target scientific community, explicitly exhibiting its discourse practices through annotated corpora. Theoretically, the salient display of rhetorical composition is expected to help students identify the patterns that are commonly used by published authors in their field and connect functional meanings with respective linguistic realizations. When students have a first draft, instructors direct them to the ‘Analyze My Writing’ Module, which gives them the opportunity to apply what they learned about disciplinary genre discourse and to practice conveying research-specific messages in their own writing.

Unlike other AWE systems, the “Analyze My Writing” Module of RWT generates metalinguistic feedback related to the rhetorical conventions of scientific argumentation. The feedback is also discipline-specific and goal-orienting, as it compares students’ drafts with articles published in their fields. RWT’s feedback takes the following forms: (1) color-coded feedback on rhetorical steps at sentence level and on communicative moves at discourse level,
and (2) goal-orienting feedback on moves and steps at discourse level (Appendix B). At sentence level, RWT initiates a dialog, telling the student what functional meaning it thinks s/he is conveying. Some feedback prompts are tentative (e.g., You are likely identifying variables here), some suggest that the sentence may be multi-functional (e.g., You may be describing instruments, but you may also be describing experimental procedures employed in your study). Some ask for clarification (e.g., It’s hard to tell... Are you describing instruments used in your study, or referring to acquiring the data, or something else?), and some indicate that the system is confused about the intended functional meaning of the sentence (e.g., Not sure what you are trying to do... Can you be more explicit?). RWT also processes the student’s draft and returns it color-coded for moves. The color codes are consistent and serve as input enhancement designed to encourage noticing. Additionally, in line with the goals of formative assessment, feedback on moves at discourse level intends to inform the students about their progress as they are working on improving their drafts, indicating how student writing may approximate the realization of moves in their discipline. The feedback on rhetorical steps at discourse level suggests that perhaps a move can be improved by revising a particular step, which may be lacking or may be under-/over-developed, at the same time acknowledging the use of steps similar to their discipline. RWT’s feedback is guiding rather than absolute. Teachers are strongly encouraged to model and emphasize self-analysis with the tool in class to ensure that students know how to engage in reflective revision and how to act upon the feedback, even if the distribution of some discourse elements is similar to the quantitative comparisons with the corpus.

**Gathering systematic evidence**

Student learning should be central to determining the types of evidence needed for a validity argument to inform the use and continuous development of RWT as an AWE complement to classroom instruction. The assumptions about the effects it might be expected to exert on students are anchored in preliminary learner-centered research. The evaluation of RWT’s prototype used by L2 writers in the targeted instructional context is reported in a book-length monograph that presents comprehensive theoretical, conceptual design, and empirical accounts for the pedagogical use of genre-based AWE (Cotos, 2014). The empirical account is based on triangulation of qualitative data (first and last drafts, screen recording during interaction with the tool, think-aloud protocols, observations, interviews, open-ended questionnaires) and quantitative data (pre-/post-tests, Likert-scale survey, frequencies of interaction with program features, automated analysis of student writing). The results obtained from researching the prototype provide a foundation for an interpretive argument for the use of RWT. Table 1 lays out several assumptions underlying respective warrants that were substantiated by evidence:

<table>
<thead>
<tr>
<th>Warrants</th>
<th>Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of the system is beneficial for learning.</td>
<td>Rhetorical feedback engages students in critical reading and enhanced cognitive activity, prompts them to identify discrepancies between intended and instantiated functional meaning, and helps them develop a meaning-oriented internal representation of their text.</td>
</tr>
<tr>
<td></td>
<td>Rhetorical feedback helps students discover and exploit the connection between steps and their linguistic realizations.</td>
</tr>
<tr>
<td></td>
<td>Focus on meaning is stimulated even when automated feedback is not entirely accurate.</td>
</tr>
</tbody>
</table>
Iterative revision and text modifications motivated by the feedback can lead to more effectively communicated rhetorical intent, improved writing quality, and genre learning gains.

Rhetorical feedback can exert positive impact on students at cognitive, pragmatic, intrinsic, and affective levels.

Students engage in modified interaction with the annotated corpus and other types of scaffolding, which can foster their learning of genre concepts, understanding of disciplinary conventions, and improvement in writing quality.

Feedback is useful for students to make decisions on revisions.

Focus on meaning fostered by rhetorical feedback helps students move from noticing a mismatch between intended and expressed meaning to external text modifications (mainly at the level of functional language use, content, and structure, and less in grammar and mechanics).

Goal-orienting feedback helps students set more substantive goals for how to operate, or what changes to make in order to improve the draft.

Feedback is relevant to writing research articles in students’ disciplines.

Rhetorical and goal-orienting feedback helps students focus on how meaning is constructed in discipline-specific genre discourse.

Feedback provides students with appropriate information to target relevant areas for revision, improvement, and learning.

Goal-orienting feedback indicating the distribution of the moves helps students notice areas that need improvement (e.g., logic in the structural arrangement of moves, considerable divergence from disciplinary patterns).

Students perceive rhetorical and goal-orienting feedback as helpful because it is task-appropriate, suitable for individual learner characteristics, and allowing the necessary degree of learner control.

Additionally, the data yielded insights that could (a) be further investigated as rebuttals in the interpretive argument, (b) help create implementation guidelines for effective pedagogical use, and (c) help further devise principles for improved design of the developing RWT. For example, data revealed that there may be instances of negative impact:

<table>
<thead>
<tr>
<th>Warrant</th>
<th>Rebuttals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of the system is beneficial for learning.</td>
<td>When students realize that certain words can help them build certain moves, they may rely mostly on lexical modifications, which can inhibit their revision strategies.</td>
</tr>
<tr>
<td>Feedback is useful for students to make decisions on revisions.</td>
<td>When students prioritize goal-orienting feedback over color-coded feedback, they may limit revision to approximating percentages and may finish revising as soon as the feedback would display percentages close to the average in their discipline.</td>
</tr>
</tbody>
</table>
When the feedback repeatedly indicates that output modifications are unsuccessful, students may become frustrated and be less motivated to continue revising.

| Feedback is relevant to writing research articles in students’ disciplines. | Students whose specific discipline is not represented in the system may see less value in revising with it. |

Guided by this framework of warranted assumptions and rebuttals, data continues to be systematically gathered to test claims pertaining to the use of RWT in the classroom, both with L1 and L2 writers, and to examine evidence that may support or refute interpretations, uses, and consequences of its automated feedback. The following is an excerpt from recently collected observation data presenting a snapshot of an L1 student’s experience with this tool, which supports some of the assumptions included in Table 1. Specifically, this excerpt illustrates how cognitive operations can be fostered by the feedback, and how focus on meaning stimulated by the feedback and the scaffolding features can lead to successful revision (note the parts underlined in Figure 3).

She copy-pastes her draft of the Introduction section into the system, clicks on ‘Analyze,’ and her draft is returned color-coded for moves. She looks at the rhetorical organization of her Introduction section. The 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 RWT thinks 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 suggests ‘You are likely stating the value of your study.’ ‘How is that different from presenting a justification?’ she wonders, and hovers over the step-level feedback 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, she thinks her meaning is not yet clearly expressed. She clicks on ‘Examples’ to see how other writers express similar claims. Here, she sees many sentences that RWT 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 (e.g., ‘strategies are needed,’ ‘must respond,’ ‘has recognized the need,’ ‘would be beneficial to develop,’ ‘innovation is required to produce’), goes back to her sentence, changes it to “Hence, an improved procedure is needed to provide a solution to this commonly encountered problem.” This time the sentence-level feedback confirms her intended functional meaning — ‘You are likely stating presenting a justification for the need to address an issue,’ so she moves on to reading the next sentence and thinking about what it does. [Student 7, observation]

Figure 3. Example of student interaction with RWT
The next example given in Figure 4 is from an L1 student’s self-reflection after using RWT for a learner-centered revision activity in class. Here, the student self-evaluates what works well in his Introduction draft, what needs improvement, and what needs to be done to improve.

```
I have a fair amount of previous research discussed and I think it is for the most part clear. I introduced the research and summarized the methods that are to come in the paper, and I think this part is done well. Move 2 by far needs the most improvement. I highlight some of the problems I address, but not all. I also don’t think that I clearly showed the gaps in current research, I need to be more explicit. I need to include more justification as to why this is important. I attempt to raise a general question, but it is also not very clear. I also feel that it is too scattered throughout the introduction. I think what is missing in move 3 is to state the value of my research. Overall I think that my paper is lacking voice. I need to be much clearer on what my goals are and why we think this is important. This is what I need to do to improve my draft:
- I need to look at the research presented and decide if some of the author prominent results should stay author prominent or if I should switch to a research prominence.
- I need to determine if the order it is presented in makes the most sense or if there is a better way.
- I need to be more explicit in stating the gaps in research, highlighting problems. I need to add the gap in 15N INEPT 2D data and 1H-15N surface species.
- I need to explain why it would be desirable to measure 15N routinely and the benefits of examining both through bond and through space.
- I also would like to see if there is a way to make a section that is more predominately the identification of the niche, rather than having it so spread out throughout the draft.
- In the final paragraph justification should be added.
- The goal of the paper also needs to be stated more clearly. [Student 12, self-reflection]
```

Figure 4. Example of student self-reflection after interaction with RWT

It has been suggested that the strongest potential benefit of AWE is its ability to support practice, metacognitive control during text production, and learning of strategies that may decrease cognitive load (Deane, 2013; Kellogg & Raulerson, 2007). An understanding of whether and how AWE can foster learners’ cognitive activity and metacognitive abilities, and whether practice in this type of technology-rich environment can improve writing competence has not yet been acquired. Nevertheless, preliminary evidence gleaned from data like the excerpts above suggest that RWT’s formative feedback with an emphasis on rhetorical conventions could be meta-cognitively supportive of the writing process and could potentially foster conditions for co-creation of meaning in light of socio-disciplinary practices. Provided in parallel with annotated corpora, such feedback could become a means for initiating an epitomized encounter with the targeted discourse community. Correspondingly, authentic corpora can be expected to create an environment where students could be more purposely emerged in the discourse of expert inhabitants of the disciplinary practice they are engaging in.

Going back to the AWA deliberations, RWT makes headway with regards to some weighty shortcomings causing the rejection of automated analysis. While not capable of truly reading for meaning, it can be used to enhance focus on communicative intent, providing a way to evaluate argumentation vis-a-vis the expectations of target scientific audiences. As one of the reviewers of this paper importantly noted, a system like RWT could be considered a customized and supported variation of procedural facilitation (Scardamalia & Bereiter, 1986), where computer algorithms can be used to generate feedback and scaffolding that would cue engagement with texts and ideas as well as provide support for reducing the information-processing demands on mental resources during the writing task. Furthermore, RWT’s analyzer is not a ‘black box’ for teachers and users; it analyzes student writing and generates output tailored to specific course objectives.
Calibrated for linguistically-realized rhetorical features of discipline-specific genre discourse, the analyzer's output emphasizes the importance of effective communication rather than subordinating meaning or depriving writing of the social and communicative dimensions. Moreover, being designed for learning and practice, this tool releases teachers of the pressure to accommodate their students’ adjustment to a scoring engine in order to ensure that they perform well on a test. The danger of scoring criteria-based assessment is replaced by RWT with a focus on guided learner’s discovery of what makes authentic genre discourse effective. Lastly, student writing, motivated by the need to enter the disciplinary discourse community for knowledge sharing, is not restricted to writing for a pre-programmed prompt.

Conclusion

Generalizing the use of AWA tools across contexts and goals is one of the most dangerous pitfalls that can impair implementation. Those involved in the progress of AWA should exercise such considerations as ecological settings (state-level assessment or in-class writing), tasks (timed content-area essay or expository writing), audience (unknown test rater or classroom teacher), and measurable skills (use of appropriate writing conventions or content-area knowledge) (Shermis, Burstein, & Apel Bursky, 201: 4). For AWE innovations, in particular, to effectively mediate writing instruction and assessment, the focus should be “on continuing to improve both the human and technological sides of the equation” (Weigle, 2013b: 50) and on the innovation that automated analysis of writing can bring, rather than on the technology itself.

Elliot and Klobucar (2013: 30) call for adopting “a unique kind of inquiry toward early developments in next generation assessments.” Although some may view it as a clash of worldviews, a unique kind of inquiry might endeavor to associate tenets from different research traditions under the argument-based validity framework. Appropriate validity evidence will substantiate the applications of automated analysis and will help foreground principled implementation. I eagerly anticipate comprehensive validity evidence describing AWE influences on learning and teaching in target instructional contexts and the “forces at play that swirl around and through this disruptive innovation” (Elliot & Klobucar, 2013: 31), which may foster or hamper successful use. Moreover, since “no developmental approach to written composition can be complete without analyzing [...] automation of certain processes and strategic control [and] real-time management of the implementation of the different components [of cognitive writing processes], whose interactiveness depends on their respective efficiency at a given level of expertise” (Alamargot & Fayol, 2009: 25), I would expect future research to investigate the effects with and of AWE systems to advance the understanding of whether and how they may support writer metacognition and foster writing development. More broadly, I foresee collaborative efforts that will produce theoretically and empirically-grounded design principles and guidelines for effective and sustainable implementation of AWA in general and of AWE in particular.

A requisite step in this direction is to leverage expertise in writing studies, cognitive psychology, artificial intelligence, learning sciences, and software design. Multi-disciplinary expertise combined with instructional and learner perspectives will provide a deeper understanding of the writing process and construct representation as valued by teachers and helpful for developing writers. While building instruction-driven systems like the genre-based RWT may seem difficult, looking into the future, pursuing such AWA transformations and
methodically studying their effects can be a highly plausible option and a significant step forward towards combining computational, humanistic, and educational perspectives.

Notes

1 Burstein et al. (2013) provide an excellent depiction of the state-of-the-art of automated scoring and evaluation of essays (including detailed descriptions of scoring systems built using supervised statistical modeling, unsupervised modeling, rule-based techniques) as well as research on their efficacy.
3 Using the Swalesian framework of moves and steps may be perceived by some as foregrounding standardization and promoting a universal template for research articles. In genre-based writing instruction, however, this framework serves as a foundation for explicit, visible pedagogy to novice research writers (see Hyland, 2007; Tardy, 2009). Equally important, it helps reduce genre complexity and facilitates their corpus-driven explorations for the discovery of similarities and differences in the genre conventions of their discipline.
4 The disciplinary corpora include a total of 900 articles published in highly-regarded journals in 30 disciplines (30 articles per discipline). It was compiled in collaboration with expert consultants, who are scholars with active research agendas in the respective disciplines. The corpus collection, move/step analysis, and annotation procedures are reported in Cotos, Huffman, & Link (in press).
Acknowledgements

The author is thankful to the guest editor of this special issue and the anonymous reviewers for their insightful comments on this manuscript. I also gratefully acknowledge the contributions of all the members of the RWT research and development team. This material is based upon work supported by the Senior Vice President and Provost/Computation Advisory Committee, Graduate College, College of Engineering, and College of Liberal Arts and Sciences at Iowa State University.

About the Author

Elena Cotos is an Assistant Professor in the Department of English, Applied Linguistics and Technology Program at Iowa State University. Her research interests include English for specific/academic purposes, corpus-based genre analysis, genre-based writing pedagogy, computer-assisted language learning and assessment, and data-driven learning. Her work has appeared in a monograph, a number of edited volumes, and in journals such as the ReCALL, CALICO, Language Testing, International Journal of Computer-Assisted Language Learning and Teaching, Journal of English for Academic Purposes, and Language Learning and Technology.
References


Appendix A

1. Annotated texts in ‘Explore Published Writing’ Module

<table>
<thead>
<tr>
<th>Discipline: Agricultural &amp; Bio-Systems Engineering</th>
<th>Section: Introduction</th>
<th>Search</th>
</tr>
</thead>
</table>

1. Introduction

Aerobic granulation is a novel self-immobilization process of microorganisms. Compared with conventional activated sludge, aerobic granules have the advantages of better settleability, stronger microbial structure, higher biomass retention, and better ability to handle toxic compounds (Adav et al., 2008; Beun et al., 1999; Liu and Tay, 2002; Morgenroth et al., 1997). The application of aerobic granules was regarded as one of the promising biotechnologies in wastewater treatment (Adav et al., 2008b). Unfortunately, aerobic granules could easily lose stability and activity during storage, which would be the serious barrier to their practical application (Adav et al., 2008a; Zheng et al., 2005). Tay et al. (2003a) and (2003b) reported that the granules became more irregular and smaller and released soluble organic materials after storage for 18 weeks. Tay et al. (2003a) also showed that granules lost about 60% metabolic activity and acetate-fed granules lost about 90% metabolic activity after 18 weeks.

Move 2: Identifying a niche: Highlighting a problem

Some storage methods were shown to improve the stability, e.g., Adav et al. (2007) proposed that storage at subfreezing temperatures (−20 degrees Celsius) was an ideal method for preserving granule stability and activity. Furthermore, the addition of toxic substances (phenol) in the storing solution was beneficial to preserve the granule stability. However, this method is not quite practical because the structure of granules would deteriorate by freezing and thawing. Although granule stability loss has been widely investigated, the information about the storage stability of aerobic granules seeded with different inocula is hardly found in the literatures so far. In most studies, aerobic granules were cultivated with activated sludge seed (Adav et al., 2008a). It is widely known that activated sludge flocs are mainly composed of cells and extracellular polymeric substances (EPS). Recently, Yu et al. (2008a, b) employed a novel EPS fractionation approach to obtain the pellets (cells) by extracting the EPS matrix. In this study, aerobic granules seeded with activated sludge flocs and pellets, respectively, were cultivated in two sequencing batch reactors (SBRs), then these two kinds of granules were both stored at −1.4 to −1 degrees Celsius for 3 weeks. The main purpose of this study was to investigate their responses to storage and explore the mechanism of stability loss. The EPS concentration variation during storage was monitored, and scanning electron microscopy (SEM) and confocal laser scanning microscopy (CLSM) were also applied to investigate the microstructure and distribution patterns of EPS and cells. Increased knowledge on this issue would further deepen our understanding of granule stability and be useful for the application of aerobic granules.


2. Move/step concordancer in ‘Explore Published Writing’ Module

Examples of "Indicating a gap"


Ecological and environmental factors that influence organisms during each stage of the dispersal process are important for understanding the movement of plant pathogens and the development of plant disease epidemics (Ashby, 1986; Madden, 1992). However, there is limited information about some of these processes, especially for the deposition of plant pathogen propagules from the atmosphere onto susceptible host plant tissues. Rainfall has been identified as an important mechanism in the spread of plant pathogen propagules (Ashby, 1986; Madden, 1992; Isard and Gage, 2001; Isard et al., 2005).


This study is the most prominent of a series of such studies in this part of Peninsular Malaysia. Current collecting indicates that inventory of the plant life on this ridge remains incomplete since the earlier listings by Ridley (1922) (21 species) and Henderson (1925) (185 species), and a later enumeration by Kiew (1982) (115 species). Many records do not distinguish between an occurrence on the quartz ridge proper and one on the adjoining (non-quartz) slopes.


Historically, construction-related activities were cited by the state as a major source of degradation to tributaries (NC DEQH, 1992), and the perception is that this has continued. While roads and highways typically make up a significant proportion of urban construction, it is unknown what the contribution of highway construction is to the total sediment export resulting from urban development. In fact, few studies have documented the effects of highway construction on the water quality of receiving streams.
Appendix B

1. Color-coded feedback on rhetorical steps at sentence level and on communicative moves at discourse level (move 1 is blue, move 2 is red, and move 3 is green).

2. Goal-orienting feedback on moves and steps at discourse level

Goal-orienting feedback is based on the analysis of the student’s draft compared to the move distribution in the disciplinary corpus. The percentages placed inside the longer bar show ranges of minimum, average, and maximum distributions of a given move. The shorter bar above points to the percent range descriptive of the distribution of each move in the student’s draft, so that the student can see whether a particular move is developed within the goal range or whether it needs more work because the draft has not enough or too much of that move.
For each move, an emoticon commends the writer on good work, and an exclamation mark indicates the steps that may need revision and improvement. For steps that need more work, the feedback may be stated as *Not enough* (or *too much*, or *lacking*) *focus on [step] compared to [discipline]*. For steps that approximate disciplinary use, the feedback states that the use of a given *[step] is very similar to [discipline]*.