A Left Turn: Automated Feedback & Activity Generation for Student Writers

Jill Burstein[1], *Beata Beigman Klebanov*[1], *Norbert Elliot* [2], *Hillary Molloy*[1]

[1] Educational Testing Service, USA
[2] University of South Florida, USA

{jburstein,bbeigmanklebanov,hmolloy}@ets.org, nelliot3@icloud.com

Abstract

Writing is key to educational and workplace success. The majority of computer-based writing support applications target K-12 students and offer standard feedback concerning the U.S.-centric, 5-paragraph, expository essay genre, which is typically taught to developing writers and used on standardized tests. This writing demo takes a "*left turn*". By "*left turn*", we mean that this demo is illustrating a new writing support concept. In contrast to most systems that provide *only* feedback, this demo illustrates how guided activities can be automatically generated as a complement to relevant writing feedback driven by natural language processing (NLP) methods. Further, the demo is intended to handle a wide range of writing genres in postsecondary academic settings. The demo feedback and activities illustrated in this paper are ideas that are intended to promote postsecondary students' engagement with writing assignments, and address writing domain knowledge that is critical for producing higher quality postsecondary writing assignments across writing genres and academic disciplines.

Index Terms: educational applications, natural language processing, writing research

Introduction

Writing is viewed as key to educational and workplace success. K-12 research has examined writing achievement and writing domain skills and knowledge [1], [2]. Writing achievement is a recognized challenge for many students in 4-year postsecondary institutions in the U.S.; yet, no parallel significant body of research exists for postsecondary students. Each year, roughly 18 million U.S. students enroll in undergraduate postsecondary education at 2- or 4-year institutions; yet, 1.7 million of these students lack the prerequisite skills to complete college-level courses, and many *drop out* [3]. Weak written communication may be a contributing factor. A disconnect has persisted for writing requirements across the K-12 and postsecondary trajectory [4], [5], [6]. Recent research has exposed a set of writing domain knowledge skills [7] for which university faculty perceived student deficiency [4].

Postsecondary institutions typically provide writing centers and other types of support resources offering students help with writing assignments across the disciplines. Some faculty in the disciplines may offer writing support beyond the discussion of disciplinary content. However, students may need help during hours when writing centers and faculty are unavailable, or may have difficulty finding time in their schedules to utilize these traditional help resources. This writing demo concept offers an alternative to traditional resources. It illustrates how guided activities can be automatically generated along with relevant writing feedback driven by natural language processing (NLP) methods. What is also different about this demo is that it begins to address a variety of writing genres across academic disciplines in postsecondary settings.

Related Work

NLP is already used to generate feedback in modern automated writing evaluation applications [8], [9]. State-of-the-art applications may offer pre-writing and planning tools, and NLP-driven, automated feedback on students' writing that highlights errors and weaknesses in a text. The feedback is typically accompanied by canned guidance. For instance, systems may provide feedback for errors in English conventions (e.g., "This sentence may be a fragment. Proofread the sentence to be sure that it has at least one independent clause with a complete subject and predicate."), and organization (e.g., "The system has not identified a thesis in your essay. A thesis is the most important sentence in an essay because it provides an overview of what you will discuss in your essay.") [10]. One system, Writing Pal, is a research system targeting K-12 that offers game-based writing activities [11] above and beyond standard feedback.

While there is evidence suggesting that existing tools can be helpful [12-18], this paper describes a demo that takes a *left turn*. Specifically, this paper presents a novel idea for writing feedback with revision activities. The "*left turn*" expands the thinking in the automated writing evaluation (AWE) technology space with regard to the ability to: 1) focus on postsecondary student writing across academic disciplines and writing genres (whereas systems have focused on K-12 writers), 2) highlight the need to draw students' attention to assignment understanding (whereas systems have focused on writing process: *planning* and *revision*), and 3) promote deep engagement and revision aimed at boosting the quality of students' writing across genres and academic disciplines through the thoughtful completion of guided activities (whereas systems have primarily focused on feedback generation and general guidance).

Demo Design

Drawing on recent work that looked at faculty perceptions related to students' weak writing quality at 4-year post-

http://dx.doi.org/10.21437/LTLT.2016-2

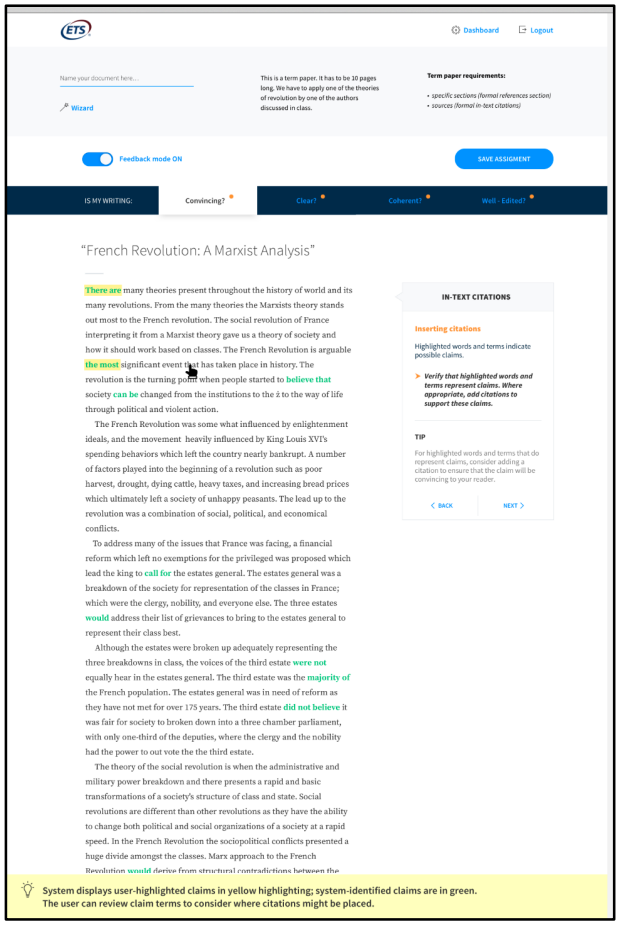

Figure 1. *Claims activity: The student highlights language indicating presence of claims; the system highlights additional terms indicating presence of claims*

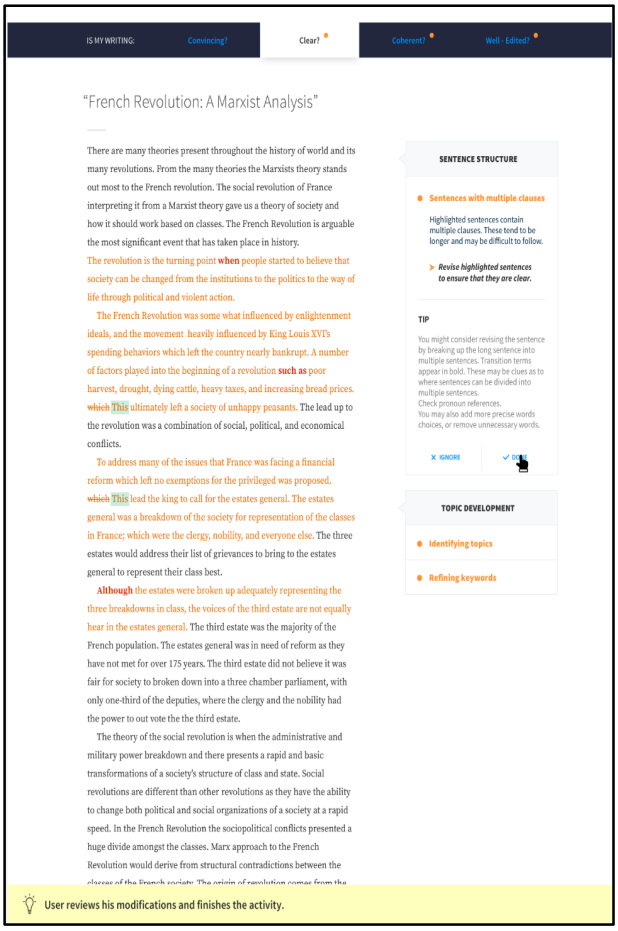

Figure 2. *Sentence revision activity: System highlights "long" sentences, and user can make revisions*

secondary institutions, the demo mock-up is organized into *writing quality categories* [4], [19]. It is intended to provide guidance and encourage students to closely review and revise critical writing construct elements aligned with a variety of writing genres across academic disciplines. The categories are, *Is my writing*: (1) *convincing?*, (2) *clear?*, (3) *coherent?*, (4) *personal?*, and (5) *well-edited*? In addition to the writing domain feedback coverage, the demo also includes an illustration of a student-system dialogue as an entry point to the system, intended to (a) help a student gauge his understanding of the assignment and (b) collect information about *self-efficacy* (one of many intra-personal factors attributed to writing achievement) [20]. In addition to writing skill domain knowledge, self-efficacy reporting could provide critical information for instructors in a real application. Such information could be used to provide analytics for instructors – e.g., an early warning that a student may be struggling due to issues aside from writing domain knowledge and may require special attention.

Once the student has worked through the full set of activities for an assignment, the demo illustrates a summary progress report to show how a system might 1) provide a summary of the student's activity engagement and progress, and 2) collect self-efficacy (or other intra-personal factor) information to assess a student's confidence after each new draft is completed. This information might be used to provide positive reinforcement, showing the student what he or she has achieved; further, (as previously mentioned) self-efficacy can be tracked along the way and used for analytics as appropriate (e.g., early warning signs of struggle, or signs of improved self-efficacy).

In the remainder of this paper through a series of screenshots, we illustrate and describe the writing support concept that targets postsecondary writers, and provides writing quality support activities and feedback. A number of existing NLP methods have been re-purposed to generate the feedback, which is illustrated in the screenshots. We provide high-level descriptions and citations to reference the NLP methods used to generate feedback displayed in the screenshots. The different activity types are motivated by the feedback. The actual demo mock-ups (as illustrated in the screenshot Figures) are renderings of actual outputs generated from the NLP methods cited for authentic writing assignments. The demo also includes a set of sketch videos created by two of the authors. An overview video describes the demo motivation, and the additional videos describe the five writing quality categories. The demo URL can be shared upon request.

3.1 Demo Feedback & Activity Sequence

The full demo mock-up illustrates the hypothetical profile of one student, "Marc S.", who attends a 4-year postsecondary

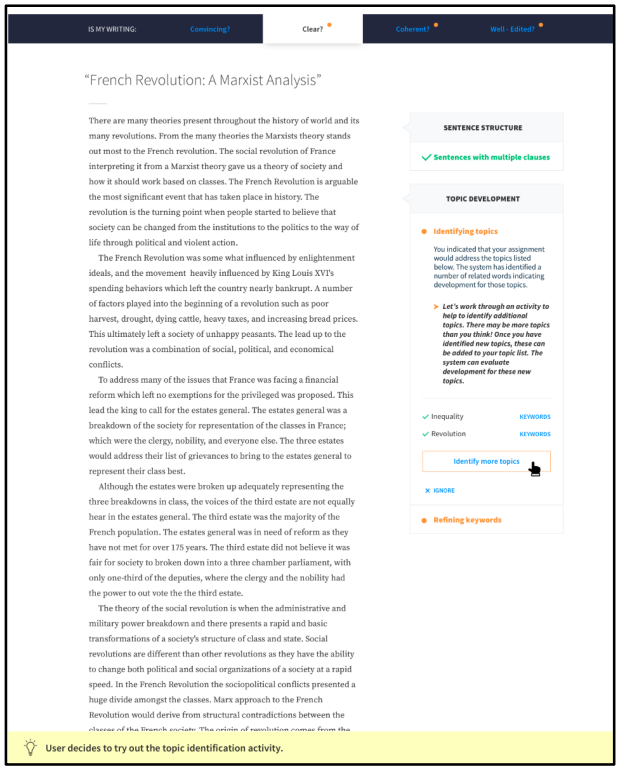

Figure 3. *Keywords identified for assignment topics: The student could click on the KEYWORDS to view the topic-related words (see Figure 7 for example keyword lists)*

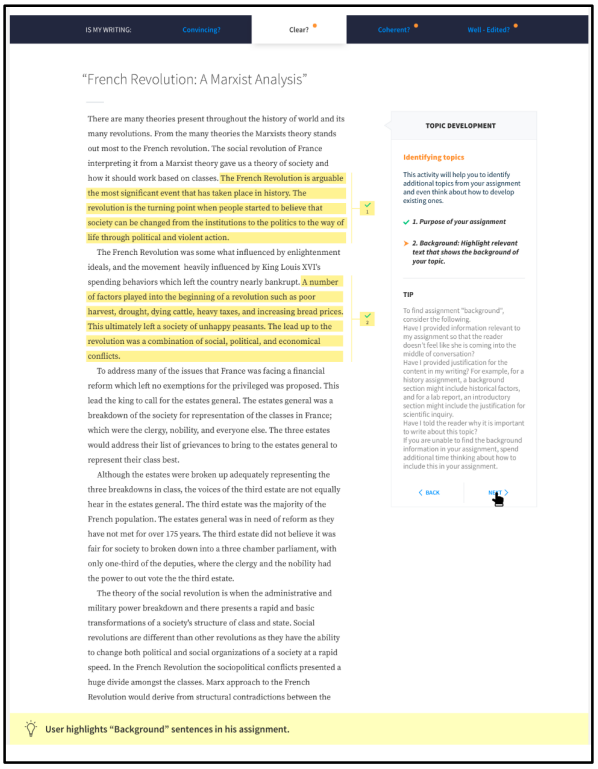

Figure 4. *Topic identification activity*: *Student highlights relevant text*

institution. The demo user is walked through "Marc S.'s" assignments. The demo illustrates his writing struggles with each assignment, the feedback and activities available at each stage, and his activity engagement and final progress. To create a coherent story in the demo, we assume that "Marc S." has written all of the assignments. The sequence of feedback and activities presented in the demo is intended to promote systematic assignment review by the student. Note that the assignments shown in the demo are authentic writing assignments. They were provided by university faculty, and come from different disciplines in the humanities, social sciences, and STEM fields. Similarly, they represent different writing genres (e.g., essays, term papers, lab reports, and research reports). In this paper, two of these assignments are used to illustrate demo feedback and activities: one comes from a political science class, and a second comes from an introductory biology class.

3.1.1 Let's Talk About Your Assignment

A recent small survey (conducted by 3 of the authors with writing instructors at American University in Washington D.C.) was intended to gain a better understanding of a) at what point in the writing process students are most likely to ask for help, and b) when students do ask for help i) what are they likely to ask for help with, and ii) what kinds of help strategies do instructors provide. Six of eight respondents indicated that "Assignment Instructions" is a common type of support for which students seek help. Instructors indicated that they *asked students questions* to verify assignment understanding, and adherence to assignment instructions.

Consistent with the survey findings, the student's point-of-entry into the demo is a "*Let's Talk About Your Assignment*" questionnaire that asks about various aspects of the assignment, including: 1) assignment genre (e.g., essay, lab report, term paper), 2) assignment description ("Provide a short description of the assignment in 1-2 sentences), 3) source requirements (e.g., formal in-text citations, or attribution), 4) section header requirements, 5) personal reflection requirement (e.g., Does the assignment require that the writer introduce his opinion, or personal experience?), 6) list of words representing expected discussion topics, and 5) self-efficacy about assignment completion (e.g., "*Do you have a good understanding of the assignment?*"). In a future system, this information could be leveraged to inform the delivery of different feedback and activities. For instance, as is indicated in Section 3.1.2, the system's knowledge of expected citations types (i.e., formal in-text citations vs. attribution) would control the type of feedback and activities offered. Feedback and activities related to "source use" would be offered only if citations or attribution were required. Once the student has completed the "*Let's Talk About Your Assignment*" questionnaire, he can continue with the set of feedback and activities described below in each of the writing quality feedback categories.

3.1.2 Is my writing convincing?

Citations are critical to credibility of claims and evidence in a text. If an assignment requires citations, the demo illustrates the presence or absence of formal in-text citations or attributions and a Works Cited section. (Note that *attributions* might be references to lecture content or other sources that do not require formal in-text citations). In the case where formal in-text citations are required, but insufficient citations are found, the

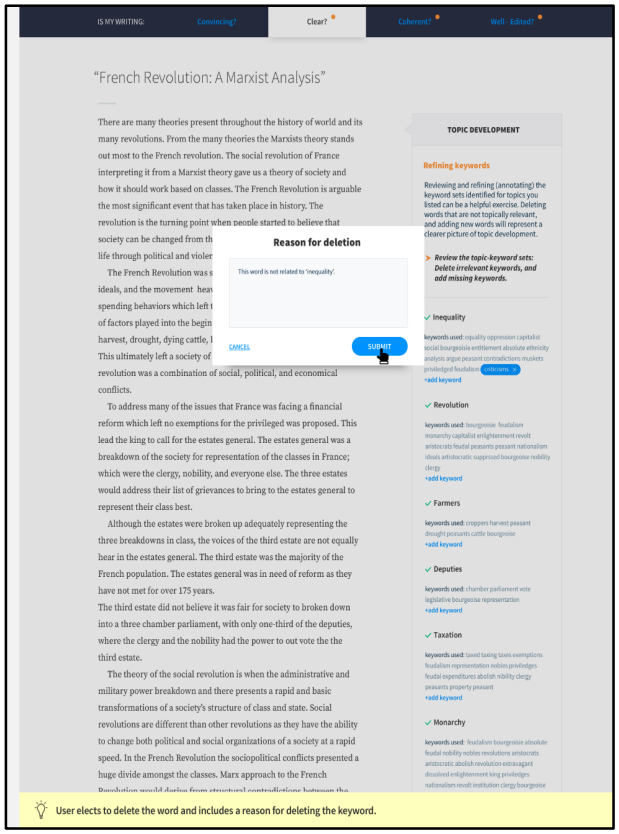

Figure 5. *Topic refinement activity: Student uses* ***'deletion' option*** *to refine topic keyword list*

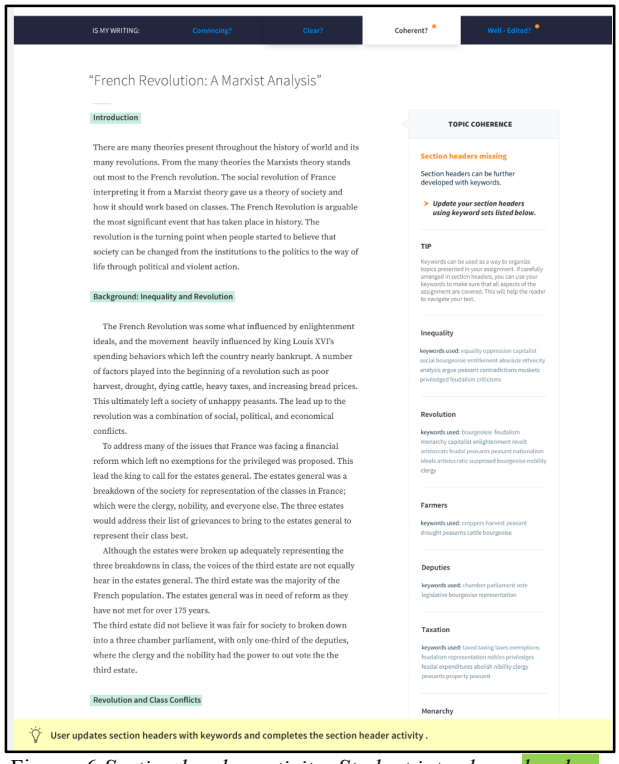

Figure. 6 *Section header activity: Student introduces headers*

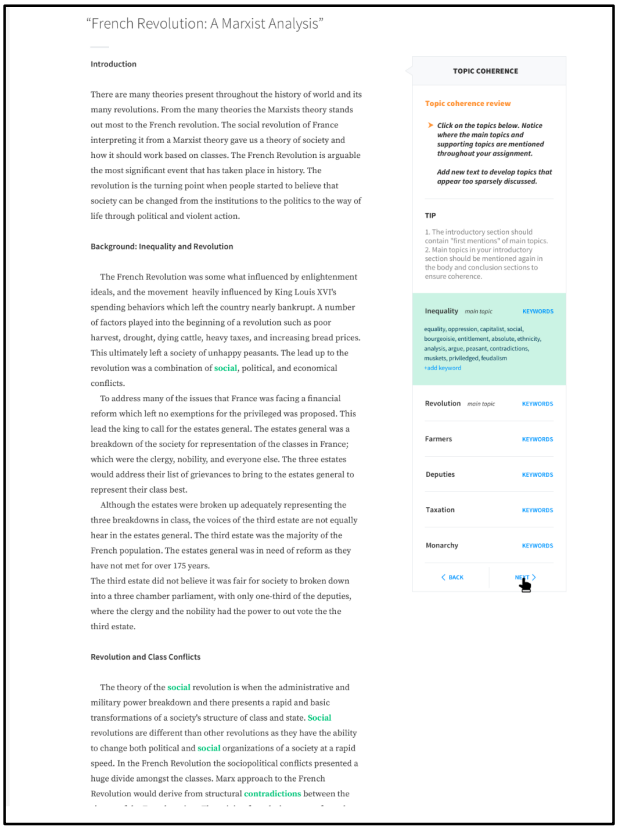

Figure 7. *Topic coherence review activity: Student reviews the distribution of main topic ('Inequality') keywords in the assignment*

demo provides an illustration of 2 activity types. These require the student to find: 1) claims and 2) evidence. The student can identify these himself, and/or he can ask for help. If the student asks for help, the demo uses lexical resources to automatically highlight language signaling "claims" (e.g., *believe that*) and automated discourse analysis to identify "evidence". System highlighting can draw attention to claims that the student may have overlooked. He is asked then to verify the automatically highlighted claims and evidence, and to consider where citations might be added to support the claims and evidence. To illustrate, we show one of the activity types: the *Claims activity* (See Fig. 1).

NLP Overview. Citations are identified using a set of regular expressions that match APA, MLA, and Chicago in-text citation styles. The "claims" terminology detection (see Fig. 1) is generated from a lexical resource designed for identifying such terminology in student essay writing. The first version of this lexicon was developed for use with an early automated essay scoring engine to delineate arguments in essays to capture content within individual text segments identified as arguments, and originally evaluated for that system [21]. The updated version of this lexical resource is used to identify claim language in the demo. This is an enhanced version of the original lexicon and contains arguing expression language based on a set of approximately 15,000 sentences from 1,000 test-taker essays in which arguing expression (claim) language (e.g., agree with, disagree with) was annotated. This updated version of the lexicon was evaluated in the context of a

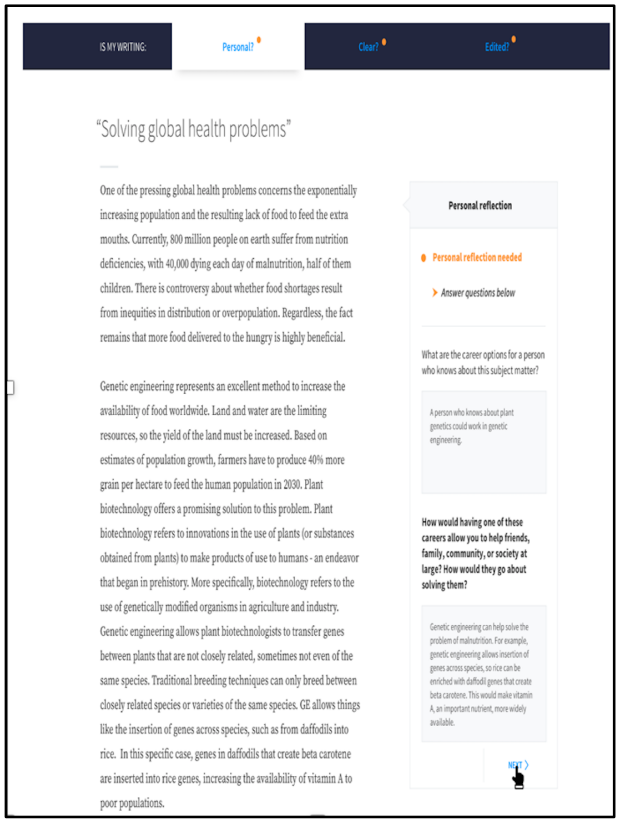

Figure. 8 *Personal Reflection Activity Excerpt: Student is presented with activity questions to elicit personal reflection. The student's responses can be included in the revision.*

discourse coherence engine that generates a coherence score for test-taker writing [22]. The evidence activity identifies text that captures detailed information that typically supports a main claim or idea. We use a discourse structure analysis system [23] used in e-rater® -- ETS' AWE system [8]. The system was designed to identify expected, *essay-based* discourse structure in test-taker and student essay writing from expository and persuasive essay writing. It identifies text segments associated with discourse elements, including thesis statements, main points, supporting details, and conclusions. To identify evidence in the demo, supporting details were highlighted.

3.1.3 Is my writing clear?

Clearly written sentences, and topic development are critical to document clarity. Feedback and activities supporting concise sentences and topic development are shown. Figure 2 shows a *sentence structure activity*, where *sentences with multiple clauses* (i.e., "long sentences") are highlighted. Since longer sentences can be confusing, the writer is asked to consider how highlighted sentences might be modified (e.g., partitioned into shorter sentences) to ensure clarity. Transition words, generated from ETS' *Criterion*® online essay evaluation system [10], are shown in boldface to provide hints as to where sentences might be partitioned. The student can revise the sentences accordingly.

NLP Overview. Regular expressions are applied to a shallow syntactic parse [21] to identify the presence of two or more dependent clauses (e.g., subordinate clauses) in the sentence [24], [25].

For *topic development*, 2 activities are offered: 1) *identifying topics* and 2) *refining topics*. These activities are motivated by the expectation that by reviewing the feedback and completing the activities, students will conduct a close read of an assignment to identify the full set of topics and to continue to develop all topics. Recall that a topic list can be entered by the student in the "*Let's Talk About Your Assignment"* step (discussed above). The list can be leveraged to generate topic-related feedback. A few different scenarios of topic development are illustrated in the demo. For instance, in a scenario where the student chooses not to enter a list of topics, the system can identify the main topic and a set of topic-related keywords. For purposes of illustration in this paper, we use a different scenario and show the 2 topic development activities.

In the first *topic identification* activity (Fig. 4), the student would see the topic list that they entered in the "*Let's Talk About Your Assignment*" section, and the topic-related keywords generated by the system (Fig. 3). Since this example assignment is a 10-page term paper, 2 topics seems like a small number, and so we now want to encourage the student to review his assignment to make sure that he has a) included critical information, and b) identified the full set of topics he has discussed, since this will be important for addressing topic coherence, later on. The student is therefore taken to an activity (Fig. 4) where he is presented with a set of 4 sub-activities adapted from Kenneth Burke's Pentad [27], originally developed to clarify motives in a literary text; we have adapted it to target motives of any writing assignment. The 4 sub-activities require the writer to identify and highlight text in the current draft that specifies: 1) *purpose* (What do I want to tell my reader?), 2) *background* (Have I provided sufficient background on the topic to the reader?), 3) *key information* (Have I provided critical evidence and important terms for my reader?), and 4) *key individuals* (Who are the critical people or organizations?). (Note that key individuals may be presented for genres where it is not relevant, e.g., a lab report.) This activity prods the student to review his assignment to ensure that this critical information is present. If information is missing, the intention is that he will be motivated to add it. In addition, this activity gives the student an opportunity to consider other main topics discussed in the assignment. Since the student has now thoroughly reviewed the assignment, he also now has an additional opportunity to update the list of 2 original topics with new topics he identified in the assignment, and consider the development of all topics. The system would then generate the set of related keywords for the additional topics (Fig. 7 illustrates that additional topics have been added). The number of keywords associated with each topic illustrates the extent to which a topic is developed. Topics associated with fewer keywords can be a cue for the writer to consider continuing to develop seemingly underdeveloped topics.

In the second *topic refinement* activity (Fig. 5), the writer is shown the full set of topics and related keyword sets identified by the system. He is prompted to review and annotate the system-generated keyword sets, by removing keywords that are not relevant, and adding new keywords that appear to be missing, while also providing rationales for each decision. This activity encourages the student's further attention to, and engagement with topics in the assignment. Further, the final, customized keyword set will illustrate development of each topic. As the student annotates, he may get a better sense of the development of each topic [28]. The writer can iterate through the topic development activities to ensure optimal topic development.

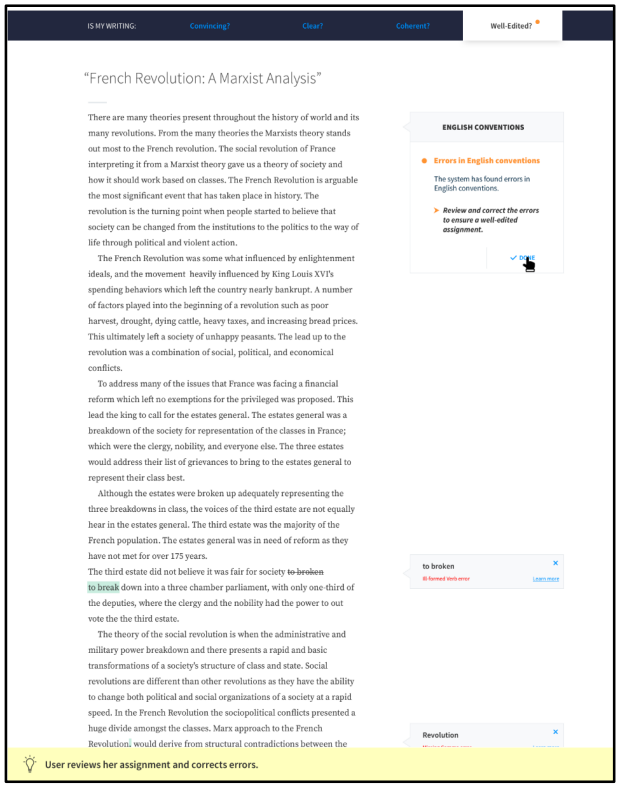

Figure 9. *Editing Feedback*

NLP Overview. Depending on the student's responses to the initial "*Let's Talk About Your Assignment"* activity, the student might or might not have provided a list of topics that he intended to address in the assignment. If no topics were provided, the system will propose a topic. For every topic, whether it is provided by the user or by the system, the system will identify a set of keywords from the essay (as rendered in Figs. 3, 5, 7). Both the topic and the keyword identification capabilities are based on the analysis of the word association graph of the essay [26]. Briefly, the vertices of the graph are all content word types in an essay; the weight on an edge quantifies the semantic relatedness between the vertices it connects. We use Pointwise Mutual Information as a measure of association [29]. Keywords for a given topic are all the *strong neighbors* of the topic word in the graph, namely, all vertices with an association above a given threshold with the topic word. In order to propose a topic (in case the user has not provided any), we select the vertex with the largest number of strong neighbors.

3.1.4 Is my writing coherent?

Coherence (flow of information) is critical to a well-written text. The demo includes coherence-motivated activities to encourage inclusion of appropriate *section headers*, and tracking of *topic prevalence (or "topic roadmap")*. *Section headers* are required for some genres requiring specific structure (e.g., lab reports), while for other genres, they are not required, but are helpful to support document clarity, e.g., term papers. For genres, such as lab reports, that require specific types of section headers, the demo provides feedback about the presence and absence of expected headers. By contrast, in Fig. 6, the mock-up illustrates a *section header activity* for a genre that does not require, but is likely to benefit from headers. In this example, the 10-page term paper had no section headers other than the title. The *section header activity* has two parts. First, the writer is prompted to consider adding major headers, (e.g., Introduction, Discussion, Conclusion). This is intended to get the student thinking about the major content "chunks". In the second part, Fig. 6 illustrates that the student is prompted to leverage the topic and related keyword sets (generated by the graph analysis method above) to consider more detailed headers: these can provide the reader with meaningful expectations about section content.

The *topic coherence review activity* (Fig. 7) presents the writer with the topic list in his assignment. He is asked to select what he believes to be the major topics. The system highlights the set of topic-related keywords in the document for those topics. (Again, the topics and keywords shown are generated from the graph centrality measure.) This provides a *roadmap* of the topic discussion throughout the assignment. The student can select sub-topics, and review the discussion roadmap for those as well. The aim of this activity is to encourage the student to review the main topic discussion *roadmaps* to ensure that these topics are discussed throughout the text. An activity "TIP" (Fig. 7) explains that if he notices a main topic is introduced early on in the assignment, but not discussed later in the text, he may want to consider revising the assignment to ensure that the topic was thoroughly discussed. Likewise, if a main topic appears only later in the assignment, he may want to consider revising the assignment to ensure that the topic was properly introduced.

3.1.5 Is my writing personal?

Some writing genres require that the writer draws on personal experience. The reflective essay used for illustration in the demo is from an introductory biology class. The assignment required that the student ask a question related to a biology unit studied in class, respond to the question using biology content knowledge, and discuss how the content related to his own life (i.e., the personal reflection piece). Some students may have trouble expressing personal reflection, and activity scaffolding could be helpful. Fig. 8 shows a novel brainstorming activity for an assignment identified as lacking required personal reflection content. An NLP method that models language use based on a corpus of writing requiring personal reflection [19] was used to determine "low personal reflection" content in the essay in Fig. 8.

NLP Overview. In previous work by two of the authors [19], writing samples from first-year students enrolled in introductory biology courses at a 4-year postsecondary institution were used to build the system for identifying presence of personal reflection language. In these samples, students were asked to pose a question related to the recently studied instructional unit, and answer the question while incorporating utility value (UV). Specifically, students needed to explain how the biology topic was related to their own life or the lives of others. Essays addressing six different biology topics are covered in the dataset (e.g., cell biology, ecology). UV and control writing assignments were scored by human raters for the level of UV articulated in each essay, on a scale of 0 (low UV) - 4 (high UV), based on the criteria related to how specific and personal the UV connection was to the writer. The experimental (UV) writing samples contained *2 types of essays* and *letters*, and the control writing samples were *summaries* wherein

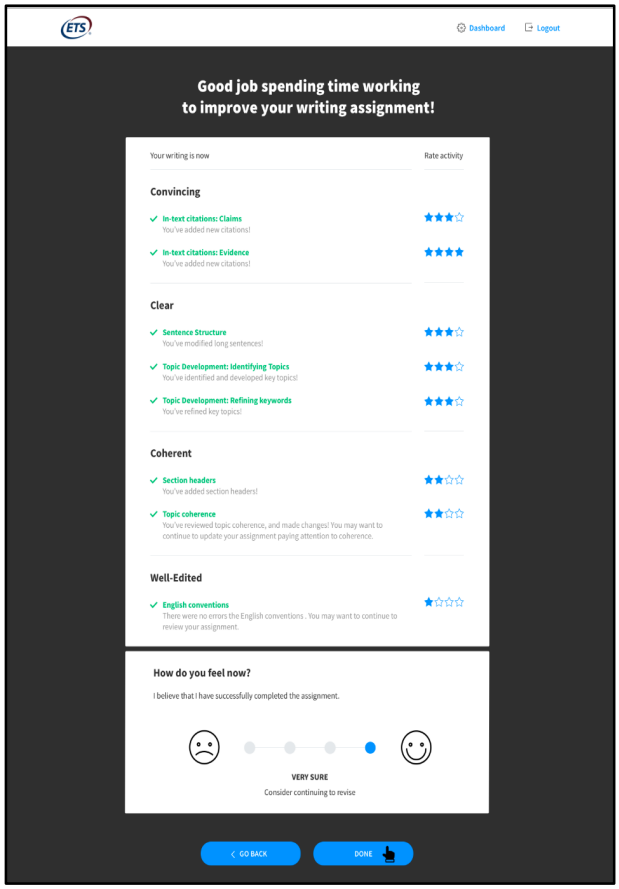

Figure 10. *Summary Progress Report*

the writer only summarized content, and UV was not explicitly required. The two types of essays represented two different writing genres, the letter a third, and the summary a fourth.

A method for predicting UV score was developed as follows, and was used on the essay in Fig 8 to predict "low UV". A number of primarily word-based language features were automatically extracted from the data and combined using the random forest regressor to predict UV score (0-4); predictions were evaluated on an independent test set. Features included the following: pronouns (e.g., 1st person singular), discourse terms expressing development (e.g., for instance), discourse terms marking epistemic stance (e.g. perhaps, inevitably), narrative elements (e.g., past tense verbs), words related to affect (e.g., nice), social processes (e.g., friends), health, and risk, and terms related to a particular genre for a particular topic (e.g., essay-ecology).

3.1.6 Is my writing well-edited?

Well-edited writing is critical to a high quality final writing product. The student is instructed to review the final draft to ensure that it is well-edited. Fig. 9 illustrates errors in English conventions that the system has automatically identified using rule-based and statistical NLP methods for error identification of grammar, word usage and mechanics errors. Errors were generated through ETS' *Criterion* system [10], which uses e-rater® -- ETS' AWE system [8]. The student is instructed to revise the errors, and is pointed to relevant resources to support revision.

3.1.7 Summary Progress Report

Figure 10 shows a summary report as presented to the student to illustrate his progress with the different activities for a given assignment. This page also collects the student's feedback on the helpfulness of each activity. This step also asks the student about self-efficacy – i.e., "confidence". The completion of activities, perceived helpfulness of activities, and self-reported self-efficacy could contribute to tracking students' writing achievement in a future system.

4. Discussion & Future Work

No one ever said that writing was easy. As we continue to develop technology-rich solutions to support student learning, it is critical that these tools offer guidance and actionable support. To facilitate high quality writing assignments, we need tools that ensure that college students 1) understand assignments, 2) engage deeply with their writing, and 3) feel confident about writing, so that students are ultimately prepared for writing in post-graduate and workplace settings. We imagine that in a future implementation of this demo, analytics could be generated to provide information about writing achievement both with regard to writing domain knowledge, skills, and abilities, as well as intra-personal factors (e.g., self-efficacy). In addition, the demo lays the groundwork for multimodal versions that could integrate spoken dialogue between the student writer and an avatar during the writing experience: from the "*Let's Talk About Your Assignment*" step through the full set of activities, and the final summary report.

In Summer 2016, the demo will be shared with university faculty. A perception survey instrument [25] has been prepared to collect faculty perceptions about the tool targeting: 1) the mock-up design and 2) the potential effectiveness of activities.

5. Acknowledgements

Many thanks to the American University administration and faculty for continued support. Thanks to Dr. Judith Harackiewicz and Dr. Charles MacArthur for their support. We would also like to acknowledge the 10clouds.com team, Magdalena Bialaczewska, Tomacz Kwiecień, and Emil Rzepiel, for their outstanding design work and mock-up development, and on-going patience. Many thanks to Diane Napolitano and Nitin Madnani for support with earlier versions of this work.

6. References

[1] V. W. Berninger, W. Nagy, and S. Beers, "Child writers' construction and reconstruction of single sentences and construction of multi-sentence texts: contributions of syntax and transcription to translation," *Reading and Writing,* vol. 24, pp. 151-182, 2011.

[2] N. G. Olinghouse, S. Graham, and A. Gillespie, "The relationship of discourse and topic knowledge to fifth-graders' writing performance," *Journal of Educational Psychology*, vol. 107, pp. 391-406, 2015.

[3] Charles A. Dana Center, Complete College America, Education Commission of the States, and Jobs for the Future. (2012). *Core Principles for Transforming Remedial Education: A Joint Statement* [Online]. Available:

http://completecollege.org/docs/CCA_joint_report-printer.pdf

[4] J. Burstein, N. Elliot, and H. Molloy, "Informing Automated Writing Evaluation Using the Lens of Genre: Two Studies," *CALICO Journal,* vol. 33, no. 1, pp. 117-141, 2016.

[5] B. Bridgeman and S. Carlson, "Survey of academic writing tasks," *Written Communication*, vol. 1, pp. 247-280, 1984.

[6] D. Melzer, *Assignments across the Curriculum: A National Study of College Writing*. Logan, UT: Utah State University Press, 2014.

[7] J. Sparks, Y. Song, W. Brantley, and L. Liu, "Assessing Written Communication in Higher Education," ETS, Princeton, NJ, Research Rep. ETS-RR-1437, 2014.

[8] J. Burstein, J. Tetreault, and N. Madnani, "The E-rater® Automated Essay Scoring System," In *Handbook for Automated Essay Scoring,* M. D. Shermis, and J. Burstein, Eds. New York, NY: Routledge, 2013, pp. 55-67.

[9] M. D. Shermis, J. C. Burstein, N. Elliot, S. Miel, and P. W. Foltz, "Automated writing evaluation: an expanding body of knowledge,". In *Handbook of writing research*, C. A. McArthur, S. Graham, and J. Fitzgerald, Eds., 2nd ed. New York, NY: Guilford, 2015, pp. 395-409.

[10] J. Burstein, M. Chodorow, and C. Leacock, "Automated Essay Evaluation: The Criterion Online Service," *AI Magazine,* vol. 25, no. 3, pp. 27-36, 2004

[11] R. D. Roscoe, L. Varner, J. Weston, S. Crossley, and D. S. McNamara, "The Writing Pal Intelligent Tutoring System: usability testing and development," *Computers and Composition*, vol. 34, pp. 39-59, 2014.

[12] M. D. Shermis, J. Burstein, and L. Bliss, "The impact of automated essay scoring on high stakes writing assessments," presented at the annual meetings of the National Council on Measurement in Education, San Diego, CA, 2004.

[13] Y. Attali, "Exploring the feedback and revision features of Criterion," presented at the National Council on Measurement in Education, San Diego, CA, 2004.

[14] A. A. Lipnevich and J. K. Smith, "I really need feedback to learn: students' perspectives on the effectiveness of the differential feedback messages," *Educational Assessment, Evaluation and Accountability*, vol. 21, no. 4, pp. 347-367, 2009.

[15] C. A. Chapelle, E. Cotos, and J. Y. Lee, "Validity arguments for diagnostic assessment using automated writing evaluation," *Language Testing,* vol. 32, pp. 385–405, 2015.

[16] J. Ranalli, S. Link, and E. Chukharev-Hudilainen, "Automated writing evaluation for formative assessment of second language writing: investigating the accuracy and usefulness of feedback as part of argument-based validation," *Educational Psychology*, pp. 1-18, 2016.

[17] L. Cassidy, K. Yee, R. Schmidt, S. Vasquez, B. Means, and A. Krumm, *Classroom Trials: A Study of Instruction with Writing Software.* Menlo Park, CA: SRI International, 2016.

[18] J. Wilson and A. Czik, "Automated essay evaluation software in English language arts classrooms: Effects on teacher feedback, student motivation, and writing quality," *Computers and Education*, vol. 100, pp. 94-109, 2016.

[19] B. Beigman Klebanov, J. Burstein, J. Harackiewicz, S. Prinski, and M. Mullholland, "Enhancing STEM Motivation through Personal and Communal Values: NLP for Assessment of Utility Value in Student Writing," In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications,* San Diego, CA, 2016, pp.199-205.

[20] C. A. MacArthur, Z. A. Philippakos, and M. Ianetta, "Self-Regulated Strategy Instruction in College Developmental Writing," *Journal of Educational Psychology,* vol. 107, no. 3, pp. 855-867, 2015.

[21] J. Burstein, K. Kukich, S. Wolff, C. Lu, M. Chodorow, L. Braden-Harder, L., and M. D. Harris, "Automated Scoring Using A Hybrid Feature Identification Technique," in *Proceedings of the Annual Meeting of the Association of Computational Linguistics*, Montreal, Canada, 1998, pp. 206-210.

[22] S. Somasundaran, J. Burstein, and M. Chodorow, "Lexical Chaining for Measuring Discourse Coherence Quality in Test-taker Essays," In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, Dublin, Ireland, 2014, pp. 950-961.

[23] J. Burstein, D. Marcu, and K. Knight, "Finding the WRITE Stuff: Automatic Identification of Discourse Structure in Student Essays," *IEEE Intelligent Systems: Special Issue on Advances in Natural Language Processing*, Vol. 18, no. 1 pp. 32-39, 2003.

[24] J. Burstein, J. Shore, J. Sabatini, B. Moulder, S. Holtzman, and T. Pedersen, "The Language Muse system: Linguistically focused instructional authoring," ETS, Princeton, NJ, Research Rep. ETS RR-12-21, 2012.

[25] N. Madnani, J. Burstein, J. Sabatini, K. Biggers, and S. Andreyev, "Language Muse™: Automated Linguistic Activity Generation for English Language Learners." in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics,* Berlin, Germany, 2016.

[26] B. Beigman Klebanov, and M. Flor, "Word Association Profiles and their Use for Automated Scoring of Essays," In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, 2013, pp. 1148-1158.

[27] K. Burke, *A Grammar of Motives,* Reprint. Berkeley, CA: University of California Press, (1945) 1969.

[28] D. Kaufer, A. Gunawardena, A. Tan, A., and A. Cheek, "Bringing Social Media to the Writing Classroom: Classroom Salon," *Journal of Business and Technical Communication*, vol. 000, no. 00, pp. 1-23, 2011.

[29] K. Church and P. Hanks, "Word association norms, mutual information, and lexicography," *Computational Linguistics,* vol. 16, no. 1, pp. 22-29, 1990.