Writing Mentor: Self-Regulated Writing Feedback for Struggling Writers

Nitin Madnani[1] **Jill Burstein**[1] **Norbert Elliot**[2] **Beata Beigman Klebanov**[1]
Diane Napolitano[1] **Slava Andreyev**[1] **Maxwell Schwartz**[1]

[1]Educational Testing Service, Princeton, NJ
[2]University of South Florida, Tampa, FL
[1]{nmadnani,jburstein,bbeigmanklebanov,dnapolitano,sandreyev,mschwartz}@ets.org
[2]nelliot3@usf.edu

Abstract

Writing Mentor™ is a freely available Google Docs add-on designed to provide feedback to struggling writers and help them improve their writing in a self-paced and self-regulated fashion. Writing Mentor uses natural language processing (NLP) methods and resources to generate feedback in terms of features that research into post-secondary struggling writers has classified as developmental (Burstein et al., 2016b). These features span many writing sub-constructs (use of sources, claims, and evidence; topic development; coherence; and knowledge of English conventions). Preliminary analysis indicates that users have a largely positive impression of Writing Mentor in terms of usability and potential impact on their writing.

1 Motivation

Low literacy is a social challenge that affects all citizens. For example, the Organization for Economic Co-operation and Development (OECD) reports that, on average, about 20% of students in OECD countries do not attain the baseline level of proficiency in reading (OECD, 2016). In the United States (US), we find literacy challenges at K-12 and post-secondary levels. The average National Assessment for Educational Progress (NAEP) reading assessment scores are only marginally proficient for 12^{th} graders in the U.S. (Musu-Gillette et al., 2017). Another important facet of the U.S. literacy challenge is the large number of English language learners (ELLs) enrolled in US K-12 schools (4.8 million in 2014–15). In post-secondary contexts, approximately 20.4 million students in Fall 2017 were expected to be enrolled in 2- and 4-year institutions. Millions of these students lack the prerequisite skills to succeed (Chen, 2016), including lack of preparation in reading and writing (Complete College America, 2012).

We describe *Writing Mentor*, an NLP-based solution to the literacy challenge that is designed to help struggling writers in 2- and 4-year colleges improve their writing at a self-regulated pace. Writing Mentor is a Google Docs add-on[1] that provides automated instructional feedback focused on four key writing skills: credibility of claims, topic development, coherence, and editing. Writing mentor builds on a large body of research in the area of automated writing evaluation (AWE) which has so far primarily been used for scoring standardized assessments (Page, 1966; Burstein et al., 1998; Attali and Burstein, 2006; Zechner et al., 2009; Bernstein et al., 2010). Burstein et al. (2017) examined relationships between NLP-derived linguistic features extracted from college writing samples and broader success indicators (such as, SAT and ACT composite and subject scores). Their findings suggested that AWE can also be useful for generating automated feedback that can help students with their writing.

Writing Mentor has been developed to provide *one-stop-shopping* for writers looking for help with academic writing. Apps such as Grammarly and LanguageTool, cater to individual users but typically focus on English conventions. Applications such as ETS' Criterion (Burstein et al., 2004) and Turnitin's Revision Assistant provide feedback beyond English conventions, but require institutional subscriptions.

[1]Freely available for use with Google Docs at https://mentormywriting.org.

Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations, pages 113–117
Santa Fe, New Mexico, USA, August 20-26, 2018.

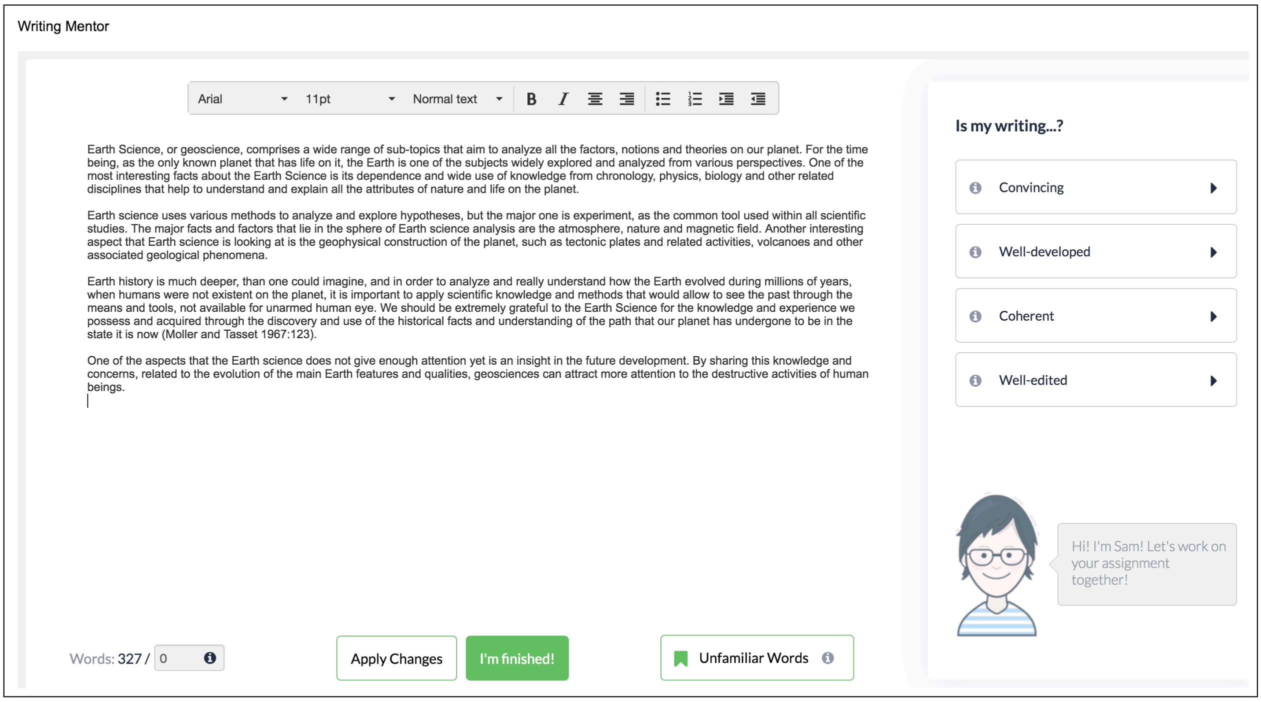


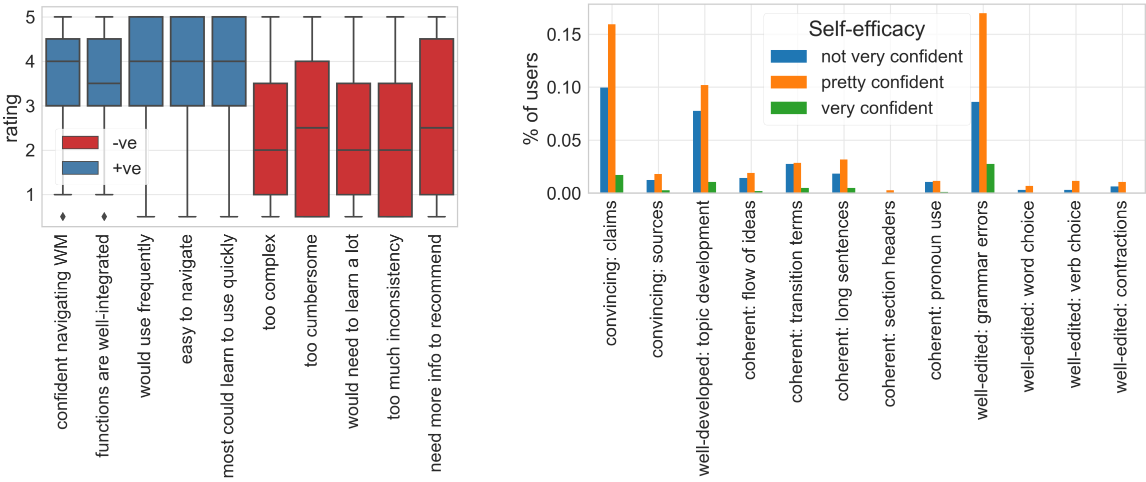
Figure 1: The Writing Mentor interface for categorized, actionable writing feedback.

Convincing	
Claims	Arguing expressions from a lexicon (Burstein et al., 1998) that contains discourse cue terms and relations (e.g., contrast, parallel, summary) and arguing expressions classified by stance (for/against) & type (hedges vs. boosters).
Sources	Rule-based system that detects in-text formal citations consistent with MLA, APA and Chicago styles.
Well-developed	
Topic Development	Detection of topics and their related word sets (Beigman Klebanov and Flor, 2013; Burstein et al., 2016a)
Coherent	
Flow of Ideas	Leverages terms in a document generated for the main topic (as identified by Topic Development above) and their related word sets.
Transition Terms	Identified using the same lexicon as in Claims above.
Long Sentences	Sentences with 1 independent clause & 1+ dependent clauses, identified using a syntactic chunker (Abney, 1996; Burstein and Chodorow, 1999)
Headers	Rule-based system using regular expressions to identify title & section headers.
Use of Anaphora	Pronouns identified using a part-of-speech tagger (Ratnaparkhi, 1996).
Well-Edited	
Grammar, Usage, & Mechanic Errors	9 automatically-detected grammar error feature types, 12 automatically-detected mechanics error feature types, and 10 automatically-detected word usage error feature types (Attali and Burstein, 2006).
Claim Verbs	Verbs denoting claims from the lexicon used in Claims above.
Word Choice	Rule-based system that detects words and expressions related to a set of 13 ‘unnecessary’ words and terms, e.g. *very*, *literally*, *a total of*.
Contractions	Identified using a part-of-speech tagger (Ratnaparkhi, 1996).

Table 1: Inventory of features provided by the NLP backend, grouped by Writing Mentor feedback types.

2 Description

Writing Mentor (WM) can be installed for free from the Google Docs add-on store. The application itself is based on a client-server model with a scalable, micro-service driven backend (Madnani et al., 2018) serving a front-end written in Google Apps Script — a JavaScript-based scripting language used for developing Google Docs extensions and add-ons. Writing Mentor was released on the Google Docs add-on store in November, 2017. Figure 1 shows the main WM interface that users interact with while writing in Google Docs. The panel on the right shows the feedback that WM provides to users – it is categorized



(a) WM Usability Perception (b) WM Feature Usage

Figure 2: Graphs showing preliminary Writing Mentor evaluations. (a) shows the distribution of ratings provided by users who chose to respond to a survey containing 10 statements (negative ones (-ve) in red, positive ones (+ve) in blue) pertaining to the usability of Writing Mentor (N=301). (b) shows the percentage of users with different levels of reported self-efficacy that preferred to spend the most time (across all documents and sessions) using each Writing Mentor feedback feature (N=1,638).

based on the writing being *Convincing* (e.g., use of claims and citation of sources), *Well-developed* (e.g., adequate topic development), *Coherent* (e.g., a good flow of ideas), and *Well-edited* (e.g., no grammatical or spelling errors). Users can receive feedback on these four aspects of their writing by clicking on the appropriate category and choosing a feedback type. For example, one could click on the *Convincing* category, and then click on "Claims" to see claims identified and highlighted in their text.

Table 1 shows an inventory of the NLP features computed by the backend and the corresponding Writing Mentor feedback type they are used for (in bold). We refer the readers to a detailed video illustrating all feedback types at `https://vimeo.com/238406360`.

3 Evaluation

We report evaluation results based on demographic and usability surveys that users voluntarily filled out and on additional information potentially correlated with the popularity of the various feedback types captured in WM's usage logs.[2]

As of May 2018, Writing Mentor has 1,960 unique users. Of these, 84% reported English being their first language. In terms of self-efficacy, 8% of all users described themselves as "very confident" writers, 40% as "pretty confident", and 52% as "not very confident". We also asked the users to rate 10 statements pertaining to WM's usability, taken from the Standard Usability Survey (Brooke, 1996). For each statement, users provided a rating on a scale of 1–5, with half point ratings also allowed. Figure 2(a) shows a distribution of the ratings users provided for each statement provided. The first five statements (in blue) represent positive impressions, e.g., "I felt confident navigating Writing Mentor" and the last five statements (in red) represent negative impressions, e.g., "I found Writing Mentor too complex".[3] The figure clearly shows that the majority of the users have largely positive impressions when it comes to the usability of Writing Mentor.

We also computed — from the WM usage logs — which of the specific WM features users spent most time in (across all of their documents and sessions) and how that varies based on the level of reported self-efficacy. Figure 2(b) shows that the three most popular features across all groups appear to be the *grammar errors* feature, followed by the *claims* feature, and then the *topic development* feature which

[2]No personally identifying information is collected by Writing Mentor. Users, documents, and sessions are assigned randomly generated IDs for logging purposes.

[3]The full text of the statements and their order the usability survey is available at `http://bit.ly/sus-usability`.

are all known to be extremely important for post-secondary writing. It also shows that, overall, users appear to be trying all WM features. From the usage logs, we also computed that approximately 25% of users return to Writing Mentor and use it again with multiple documents. Repeat use likely indicates that a user is actually benefiting from using Writing Mentor.

4 Conclusion

We described Writing Mentor – a freely available Google Docs add-on that can help struggling post-secondary writers improve their academic writing by providing automatically generated, categorized, and actionable feedback on various aspects of their writing using NLP resources and techniques. We conducted some preliminary evaluations and observed that users have a largely positive impression of Writing Mentor's usability, they are spending time using Writing Mentor features that are known to be important for post-secondary academic writing, and that many of them are returning to use Writing Mentor for multiple documents.

References

Steven Abney. 1996. Part-of-Speech Tagging and Partial Parsing. In *Corpus-Based Methods in Language and Speech*, pages 118–136.

Yigal Attali and Jill Burstein. 2006. Automated essay scoring with e-rater [r] v. 2. *Journal of Technology, Learning, and Assessment*, 4(3).

Beata Beigman Klebanov and Michael Flor. 2013. Word Association Profiles and their Use for Automated Scoring of Essays. In *Proceedings of ACL*, pages 1148–1158.

Jared Bernstein, A. Van Moere, and Jian Cheng. 2010. Validating Automated Speaking Tests. *Language Testing*, 27(3):355–377.

John Brooke. 1996. SUS - A Quick and Dirty Usability Scale. *Usability evaluation in industry*, 189(194):4–7.

Jill Burstein and Martin Chodorow. 1999. Automated Essay Scoring for Nonnative English Speakers. In *Proceedings of a Symposium on Computer Mediated Language Assessment and Evaluation in Natural Language Processing*, pages 68–75.

Jill Burstein, Karen Kukich, Susanne Wolff, Chi Lu, Martin Chodorow, Lisa Braden-Harder, and Mary Dee Harris. 1998. Automated Scoring Using A Hybrid Feature Identification Technique. In *Proceedings of ACL/COLING*, pages 206–210.

Jill Burstein, Martin Chodorow, and Claudia Leacock. 2004. Automated Essay Evaluation: The Criterion Online Writing Service. *AI Magazine*, 25(3):27.

Jill Burstein, B Beigman Klebanov, Norbert Elliot, and Hillary Molloy. 2016a. A Left Turn: Automated Feedback & Activity Generation for Student Writers. In *Proceedings of the 3rd Language Teaching, Language & Technology Workshop*.

Jill Burstein, Norbert Elliot, and Hillary Molloy. 2016b. Informing Automated Writing Evaluation Using the Lens of Genre: Two Studies. *CALICO Journal*, 33(1).

Jill Burstein, Dan McCaffrey, Beata Beigman Klebanov, and Guangming Ling. 2017. Exploring Relationships Between Writing & Broader Outcomes With Automated Writing Evaluation. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 101–108, Copenhagen, Denmark, September.

Xianglei Chen. 2016. Remedial Coursetaking at U.S. Public 2- and 4-Year Institutions: Scope, Experiences, and Outcomes. `https://nces.ed.gov/pubs2016/2016405.pdf`. [Online; accessed 01-May-2018].

Complete College America. 2012. Remediation: Higher Education's Bridge to Nowhere. `http://files.eric.ed.gov/fulltext/ED536825.pdf`. [Online; accessed 01-May-2018].

Nitin Madnani, Aoife Cahill, Daniel Blanchard, Slava Andreyev, Diane Napolitano, Binod Gyawali, Michael Heilman, Chong Min Lee, Chee Wee Leong, Matthew Mulholland, and Brian Riordan. 2018. A Robust Microservice Architecture for Scaling Automated Scoring Applications. *ETS Research Report Series*, 2018(1).

Lauren Musu-Gillette, Cristobal De Brey, Joel McFarland, William Hussar, William Sonnenberg, and Sidney Wilkinson-Flicker. 2017. Status and Trends in the Education of Racial and Ethnic Groups 2017. `https://nces.ed.gov/pubs2017/2017051.pdf`. [Online; accessed 01-May-2018].

OECD. 2016. PISA 2015 Results in Focus. `https://www.oecd-ilibrary.org/content/paper/aa9237e6-en`. [Online; accessed 01-May-2018].

Ellis B. Page. 1966. The Imminence of ... Grading Essays by Computer. *The Phi Delta Kappan*, 47(5):238–243.

Adwait Ratnaparkhi. 1996. A Maximum Entropy Model for Part-of-Speech Tagging. In *Conference on Empirical Methods in Natural Language Processing*.

Klaus Zechner, Derrick Higgins, Xiaoming Xi, and David M. Williamson. 2009. Automatic Scoring of Non-native Spontaneous Speech in Tests of Spoken English. *Speech Communication*, 51(10):883–895.