Improving student writing through automated formative assessment: Practices and results

Peter W. Foltz
Pearson and University of Colorado, Boulder
Peter.Foltz@pearson.com

Abstract

Writing practice is a key component to building mature language skills. However, because hand scoring of writing is time consuming, it is often not possible to provide rapid individualized feedback to students to maximize their writing and language skills. This paper describes the development, use, and results from an implementation of a grade school level formative writing environment which provides accurate, instant automated feedback to student writers of English essays. The system assesses writing across a range of skills and provides feedback to students on aspects such as grammar, content, and organization. Students are further provided support for planning and revising, as well as writing summaries from readings. Instructors are provided with rapid summaries of student learning growth. This paper describes results from studies that show improvement in student writing and language skills for native English speakers and English language learners. Implications are also discussed for technologies for supporting writing English language learners.

Keywords: Formative Writing Practice, English Language Learning, Automated Scoring, Learning to write.

Supporting and improving writing through formative feedback with WriteToLearn

Learning to read and write well partially comes through doing a lot of reading and writing. However, while time on task is typically a strong predictor of performance gains in reading and writing, receiving timely feedback is critical (e.g., Black & William, 1998, Landauer et al., 2008). Recent meta-analyses of research studies have identified formative writing practices that can best improve student reading and writing performance. The studies, *Reading Next* (Biancarosa and Snow, 2004), *Writing Next* (Graham and Perin, 2007), *Writing to Read* (Graham and Hebert, 2010), and *Informing Writing* (Graham, Harris & Hebert, 2011), all indicate a range of practices which can be applied in formative writing, as well as the effect size on student learning. A number of the practices are amenable to incorporation into applications for formative writing practice. These practices include:

- Teaching students strategies for planning, revising, and editing their compositions (effect size=0.82)
- Teaching students how to assess their own writing (effect size=.46)
- Explicitly and systematically teaching students how to summarize texts (effect size= 0.82)
- Providing feedback (effect size =.77)
- Monitoring students' writing progress e(ffect size=.24)
- Collecting multiple samples of student writing for teachers

Communicating information through writing is considered one of the key 21st Century skills and has been incorporated as a critical component in many national standards (e.g., Ananiadou & Claro, 2009; OECD, 2013). For example, the U.S. Common Core State Standards require students to develop more rigorous writing skills with a stronger emphasis on the ability to synthesize and summarize informational text, formulate arguments, as well as respond appropriately to source documents. This puts greater emphasis on writing for a purpose, with students linking ideas to texts and argumentation. Thus, what is needed are improved ways of having students write frequently, encountering topics and texts at the appropriate level of complexity across the curriculum, while receiving timely feedback and having instructors able to monitor the writing over time and obtaining reliable, valid, and timely data on the students.

Automated scoring for formative writing practice

A key limitation of implementing increased writing in the classroom is the amount of time that instructors would need to review, edit, and comment on student writing. Typically a teacher can only review student essays after a class period and even then, student may not receive the feedback on their writing a day or a week later. This limits learning since students learn best when receiving timely feedback (e.g., Anderson et al., 1990). Automated scoring of student writing provides a means to give students instant specific, immediate, feedback that addresses both the content of their writing, as well as surface-level features As such, it gives students increased opportunities to practice writing skills and use the feedback to improve writing.

Automated assessment of writing has become increasingly accepted with multiple systems available for implementing the scoring of writing (e.g., Shermis & Burstein, 2013).

Studies of these systems have shown that the scoring of such systems can be as accurate as human scorers (e.g., Burstein, et al., 2004; Landauer, Laham & Foltz, 2001; Shermis & Hamner, 2011), can score on multiple traits of writing (e.g., Foltz et al., 2013), and can be used for feedback on content (Foltz, Gilliam, & Kendall, 2000). It is also becoming more widely used for formative assessment.

In this paper, we describe the implementation of WriteToLearn, a web-based formative tool designed to improve writing through incorporating the writing practices and strategies. WriteToLearn's automated writing scoring is based on an implementation of the Intelligent Essay Assessor (IEA). IEA is trained to associate extracted features from each essay to scores that are assigned by human scorers. A machine learning-based approach is used to determine the optimal set of features and the weights for each of the features to best model the scores for each essay. From these comparisons, a prompt and trait-specific scoring model is derived to predict the scores that the same scorers would assign to any new responses. Based on this scoring model, new essays can be immediately scored by analysis of the features weighted according to the scoring model (see Landauer, Laham, & Foltz, 2001; Foltz et al., 2013). IEA further provides advisories, detecting such things as off-topic essays, common forms of trying to enhance the essays in construct irrelevant ways (e.g., use of inappropriate words, repetition), and plagiarism.

WritetoLearn is implemented to allow students to write, receive detailed feedback, and revise essays and summaries in order to improve skills rather than just receive snapshot measures of writing performance. The software consists of two different types of writing tasks, essay writing and summary writing. The essay writing portion of WriteToLearn provides detailed feedback on students writing, as well as information on how to improve their essays. The summarization portion of WriteToLearn lets students practice summary writing across diverse content areas and provides feedback on how well the student has covered the content of each major section of the document that the student has read. The read, write, and revise cycle encourages the students to re-read and re-express those parts of the text that they have not as well understood. Finally, WriteToLearn is designed as a tool for teachers, allowing them to have realtime monitoring of students' performance, to provide feedback to students, and to maintain a portfolio of student writing.

Formative Essay Writing in WriteToLearn

In writing to specific prompt, students receive feedback using standard traits and rubrics used to score state writing exams. The rubrics are typically holistic scores on a 4- or 6-point scale as well as specific trait scores, such as: ideas, organization, conventions, word choice, and sentence fluency. In addition, redundancy, grammar and spelling errors are flagged. Figure 1 below shows the system's scoring of a persuasive prompt. The bars show the scores achieved on each of the writing traits, while the triangles above the bars show the student's performance on the prior draft of the essay. Teacher comments on drafts are also displayed within the student interface

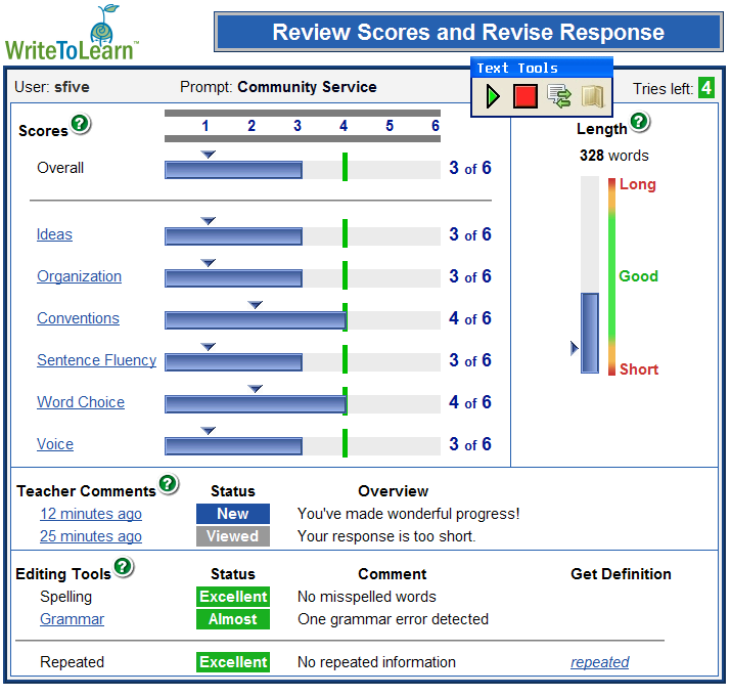

Figure 1. Sample essay feedback to a persuasive writing prompt.

Within the essay feedback, students can click on the links associated with each trait to receive feedback tied to their specific trait score. Clicking opens up a separate window with information about what students can do to their essay to improve writing on that trait. It further includes access to a sample essay (on a different topic, but same genre) that provides examples of good and poor writing for that specific trait and for the score that the student received. The student can also view better examples the essay to help them determine how they could re-write them to achieve a better score (see Figure 2). Based on this feedback, a student can go back and revise and resubmit the essay.

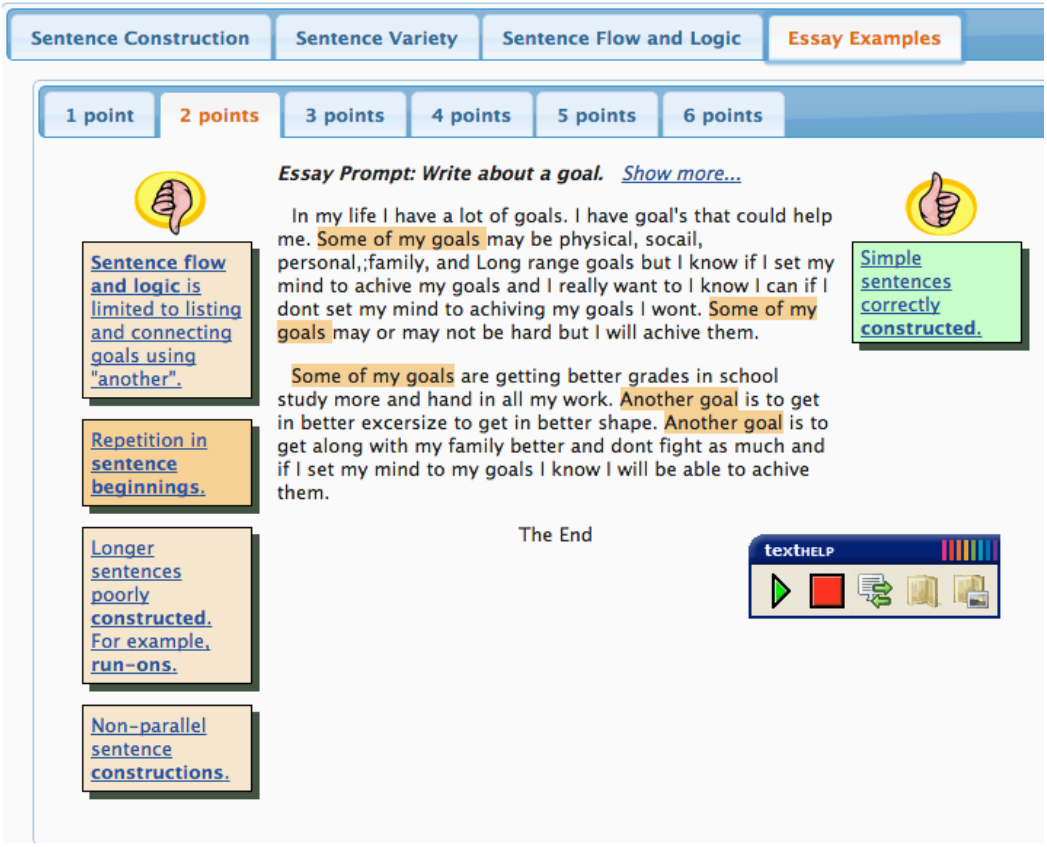

Figure 2. Sample of training on sentence constructions

Summary writing

Summary Writing within WriteToLearn is used as a means to help train reading comprehension, as well as learning to write. In the summary writing component, students are assigned a text from over a thousand readings in science, social studies, history and general interest that range from fourth grade up to high school. After reading the text, the student is asked to write a summary of what they read and then click submit to receive feedback. An example of student feedback is shown in Figure 3. It includes an assessment of how well the student covered the content in each major section of the reading, hints for how to improve content coverage in a particular section, and feedback on length, unimportant content, redundant content, and direct copying from the original text. Scoring is accomplished by analyzing both the passage sections and summary for their holistic meanings, not by looking for particular key words (see Foltz, Gilliam & Kendall, 2000). For sections on which students do not perform well on, students are able to click on "hints" to be asked questions about the text which may help them think more deeply about particular parts of the text. They are also able to click on section headings which will bring up the original text for them to re-read so that they may improve their summary on subsequent revisions.

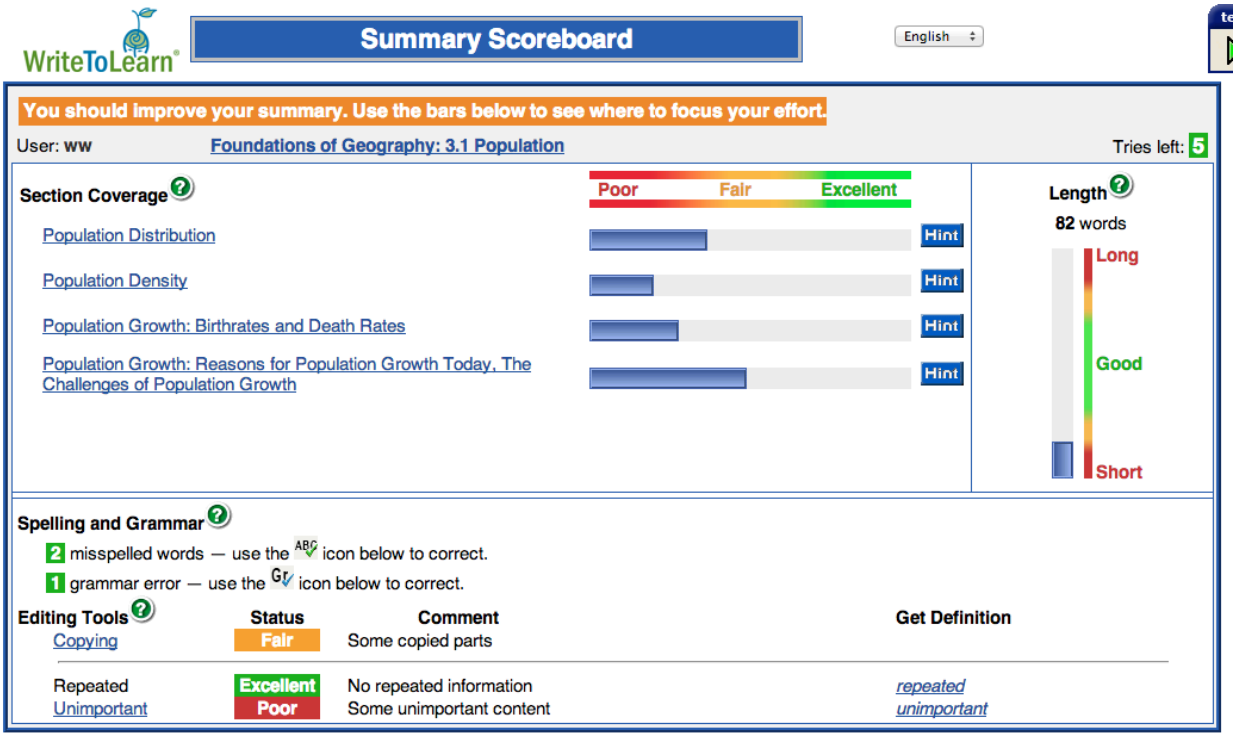

Figure 3. Sample feedback on a student summary of a chapter on population growth.

Effectiveness of automated formative writing

A number of studies have been performed to examine the effects of the summary writing and prompt writing using WriteToLearn within the classroom.

For summary writing, studies have examined both reading comprehension and writing skill using the summarizing feedback. In one study, four eight grade classes received four weeks of training on summarization. Half of the students received automated summary feedback, while the other half control group who received the same training but did their summary writing on word processors, without feedback (Franzke, et al., 2005)). Students receiving feedback improved their summary writing by an overall effect size of $d = 0.9$ compared to the control students. (An effect size of 1.0 corresponds to approximately a one-grade difference, e.g., from eight to ninth grade.). Mid-level students (those scoring at the fiftieth percentile) improved their writing performance with more difficult materials to the eighty-second percentile. The low- and medium-ability students (those in the lower 75 percent of the distribution) showed an effect size to $d = 1.5$ for the most difficult materials.

Another study conducted a large evaluation of 2,851 students in grades 5-9 across nine Colorado schools districts over two years(Caccamise, et al., 2009). Classes of students were assigned to either use the summarization tools of WriteToLearn or to receive traditional teacher-provided summarization instruction. Each group summarized approximately six texts in a year. Results showed that students receiving automated summarization feedback were superior to the control groups over both years of the study. Improvement in summarization was also highly related to the number of texts a student studied during the year, as well as the amount of time students spent using the tool. While the results showed that the tools supported summarization

skills, tests using the Test of Reading Comprehension (TORC) also showed significant effects of improvement based on the number of texts that students summarized.

In evaluating students writing in response to prompts in WriteToLearn, one large study examined a state-wide implementation of WriteToLearn. Based on an analysis of 21,137 students writing to 72,051 assignments (an average of almost four assignments per student), students generated over a quarter million essays in the course of four months. For each submission, students received feedback and scores on their overall essay quality, as well on six different writing traits. An analysis of the administration (Foltz, Lochbaum & Rosenstein, 2011; Foltz & Rosenstein, 2013) showed that, on average, students would revise an assignment four times; more revision practice than could or would occur in a conventional classroom with teacher grading. Over the revisions, students on average improved their scores by one grade point out of a total of six. The results further showed not just improvement for basic writing skills such as grammar, but also for traits like ideas, and organization.

Writing Support for English Language Learners

While the above results indicate that automated scoring embedded within formative writing systems can be effective for native English language speakers, there are critical challenges for teaching writing for learners who encounter writing in English as a second or foreign language. When approaching writing for English language learners, first, the focus may initially be on lower level writing skills such as grammar, mechanics (e.g., spelling and punctuation) and word choice rather than higher level features such as organization and flow of ideas. Second, additional support must often be provided to scaffold the writing process (e.g., Bradley & Bradley 2004). This can include providing additional writing support in generating the written text, as well as more support on using and understanding academic vocabulary. Finally, support must be provided that allows bridging of the students first language to the writing in English.

These approaches have been addressed through implementing: 1) revised writing rubrics to support language learners, 2) support tools for language learning, and 3) incorporating translation and visual representations of words. The approach to incorporating these into formative writing are described below.

Rubrics for English language learners.

As described above, language learners often need greater focus on the acquisition of basic writing skills and therefore writing rubrics must provide feedback on those skills. To this end, we have implemented additional rubrics with modified writing traits that can be used in place of the original six writing traits. The writing traits and rubrics are shown in Figure 4. The four traits focus on fundamental writing skills, including Sentence correctness, Language usage, Mechanics, and the Presentation of ideas. Essays are automatically scored on a four point scale for each of these traits, allowing students to determine the types of errors upon which they can focus to improve their writing performance.

	1	2	3	4
Sentence Correctness	• Little to no evidence of the ability to write correct sentences. • Numerous sentence errors occur. • Some errors are serious. • May be multiple errors in some sentences. • Errors may impede meaning.	• Limited evidence of the ability to write correct sentences. • Sentence errors occur in some sentences. • Some errors may be serious. • Some fixes require rewriting. • Errors may be distracting and interfere with meaning.	• Evidence of the ability to write correct simple sentences. • Uses mostly simple sentences which are mostly correct. *OR* • Uses mainly longer sentences but some may contain minor errors. • May have some errors in longer sentences. • Most errors do not interfere with sentence sense or meaning.	• Evidence of the ability to write correct sentences. • Sentences are mostly correct. • Includes longer, more complicated sentences. • May have a very few minor errors in longer sentences. • Errors do not interfere with sentence sense.
Language Usage	• Numerous errors occur in grammar and word usage. • Some errors are serious. • Multiple errors may occur in some sentences. • Errors may impede understanding.	• Grammar and word usage errors occur. • May have some errors in each category or severe errors in one category. • Some fixes require rewriting. • Errors may be distracting.	• Grammar and word usage errors may occur. • Most errors are minor and easily fixed. • Most fixes are of the one word variety.	• Grammar and word usage is mostly correct. • May have a very few minor errors.
Mechanics	• Punctuation, capitalization and/or spelling errors occur. • May have many basic errors in a short essay *OR* • Many different kinds of errors in a longer essay. • Errors may impede understanding.	• Punctuation, capitalization and/or spelling errors occur. • May have some errors in each category or • A density of errors in one category. • Errors may be distracting.	• Some punctuation, capitalization and/or spelling errors may occur, but are minor and easily fixed. • Many sentences have no errors.	• Punctuation, capitalization and spelling are mostly correct. • Most sentences have no errors.
Presentation of Ideas	• Little to no evidence of a plan. • May have introductory sentence that is like an answer to the "question". • May use "and" or "also" to list ideas or information. • May use a transition or two but without real meaning or effect. • Focus may be limited or severely disrupted.	• Limited evidence of an overall plan, but some clear logic to the presentation of ideas information. • May have beginning or ending. • Provides some connections or transitions between sentences and ideas. • Focus may be disrupted.	• Some evidence of a plan for presenting ideas and information. • Identifiable beginning and ending. • Some transitions used to introduce and order ideas and information. • Focus mostly sustained.	• Strong evidence of a plan for presenting ideas and information. • Introduction conclusion. Transitions used to introduce and order ideas and information. • Sustained focus and overall cohesiveness.

Figure 4. Writing Rubrics for English Language Learners.

Scaffolding learning through support tools.

To support the English language learner in writing, additional tools are provided to allow easier generation of language. Each of these is provided as options which an instructor can turn on if they feel that students need them. One aspect that can help students in reviewing their own writing is to hear it read back to them. Therefore, a text-to-speech system is provided to allow them to catch errors that may not be seen visually, but can be heard when it is read. Because generation of text is often difficult, a second aspect provided is a word prediction tool. This tool predicts words as students type and allows students to select correct word choices. This simplifies some of the work required by students in word generation and spelling allowing them to focus more on other aspects of writing. Finally, within the reading passages, students can highlight main and supporting ideas before generating summaries, helping them to build their comprehension skills.

Language Support

Although students may be writing in English, they may not be able to understand all the feedback if it is only provided in English and uses unfamiliar terms. Thus, both instruction and feedback can be provided in other languages such as Spanish and simplified Chinese. As students are writing or when reading passages, they may also encounter or create words they do not know. Therefore and embedded dictionary is provided which can perform cross-language translation at the word level, as well as a picture dictionary providing graphic representations of words.

Conclusions

Overall, the results from research studies and operational implementations of WriteToLearn show that it serves as an effective means to provides students with learning to write responses to narrative, expository, descriptive, and persuasive prompts as well as to read texts and write summaries of the texts in order to build reading comprehension. By incorporating automated scoring, WriteToLearn is able to provide rapid, accurate feedback, detecting the kinds of errors made by students so that the feedback is scaffolded based on the students' writing performance. For English Language Learners, learning to write must be supported in a way to allow students to focus more on the process of writing and allow scaffolding of different language components that may be needed to help in writing. While supporting student learning, automated scoring in formative writing tools also serves as a tool for teachers, allowing them to assign more writing, while giving them the ability to continue to read and grade essays.

References

Ananiadou, K., & Claro, M. (2009). 21st century skills and competences for new millennium learners in OECD countries, OECD Education Working Papers, 41.

Biancarosa, C., & Snow, C. E. (2006). Reading next: A vision for action and research in middle and high school literacy. A report to Carnegie Corporation of New York (2nd ed). Washington, DC: Alliance for Excellent Education.

Black, P., & Wiliam, D. (1998). Inside the Black Box: Raising Standards Through Classroom Assessment [Electronic version]. *Phi Delta Kappan, 80*, 139-148. Retrieved March 3, 2014, from http://www.pdkintl.org/kappan/kbla9810.htm

Bradley, K. S., & Bradley, J. A. (2004). The Internet TESL Journal, Vol. X, No. 5, May 2004 http://iteslj.org/

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The Criterion Online writing service. *AI Magazine, 25*(3), 27-36.

Caccamise, D.J., Snyder, L., Allen, C., Oliver, W., DeHart, M., Kintsch, E., et al. (2009). Teaching comprehension via technology-driven tools: a large scale scale-up of Summary Street. Report to *IES.*

Foltz, P. W., Gilliam, S., & Kendall, S. (2000). Supporting content-based feedback in online writing evaluation with LSA. Interactive Learning Environments, 8, 111–129.

Franzke, M., Kintsch, E., Caccamise, D., Johnson, N., & Dooley, S. (2005). Summary Street: Computer support for comprehension and writing. *Journal of Educational Computing Research,* 33, 53–80.

Graham, S., & Harris, K. R. (2005). Improving the writing performance of young struggling writers: Theoretical and programmatic research from the Center on Accelerating Student Learning. *Journal of Special Education, 39*, 19–33.

Graham, S., & Perin, D. (2007). Writing next: Effective strategies to improve writing of adolescents in middle and high schools – *A report to Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education.

Graham, S., Harris, K., & Hebert, M. A. (2011). Informing writing: The benefits of formative assessment. *A Carnegie Corporation Time to Act report*. Washington, DC: Alliance for Excellent Education.

Foltz, P. W., Kintsch, W,. & Landauer, T. K. (1998). The measurement of textual coherence with Latent Semantic Analysis. *Discourse Processes, 25*(2&3), 285-307.

Foltz, P. W., Laham, D., & Landauer, T. K. (1999). The Intelligent Essay Assessor: Applications to Educational Technology. *Interactive Multimedia Education Journal of Computer Enhanced Learning*. *1,* (2).

Foltz, P. W., Streeter, L. A., Lochbaum, K. E., & Landauer, T. K (2013). Implementation and applications of the Intelligent Essay Assessor. *Handbook of Automated Essay Evaluation*, M. Shermis & J. Burstein, (Eds.). pp. 68-88. Routledge, NY. NY.

Landauer, T.K., Lochbaum, K.E., & Dooley, S. (2009). A new formative assessment technology for reading and writing. *Theory into Practice*, 48. 44-52

Landauer, T. K., Laham, D. & Foltz, P. W. (2001). Automated essay scoring. *IEEE Intelligent Systems*. September/October.

Shermis, M. and Hamner. B. (2012). Contrasting state-of-the-art automated scoring of essays: Analysis. In *Paper presented at Annual Meeting of the National Council on Measurement in Education*, Vancouver, Canada, April.